Columbia University
Department of Medical Informatics
April 17, 2003

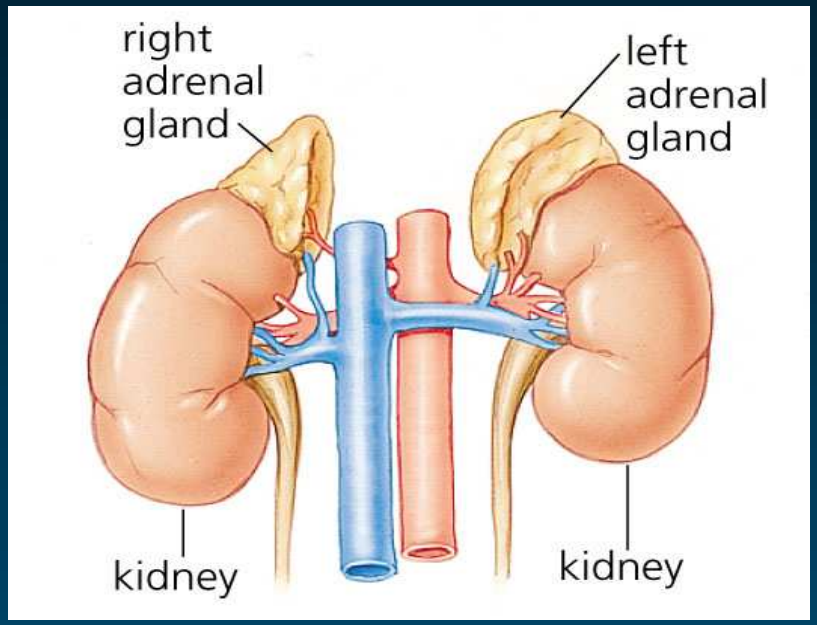
The Unified Medical Language System: Between Terminology and Ontology

Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA

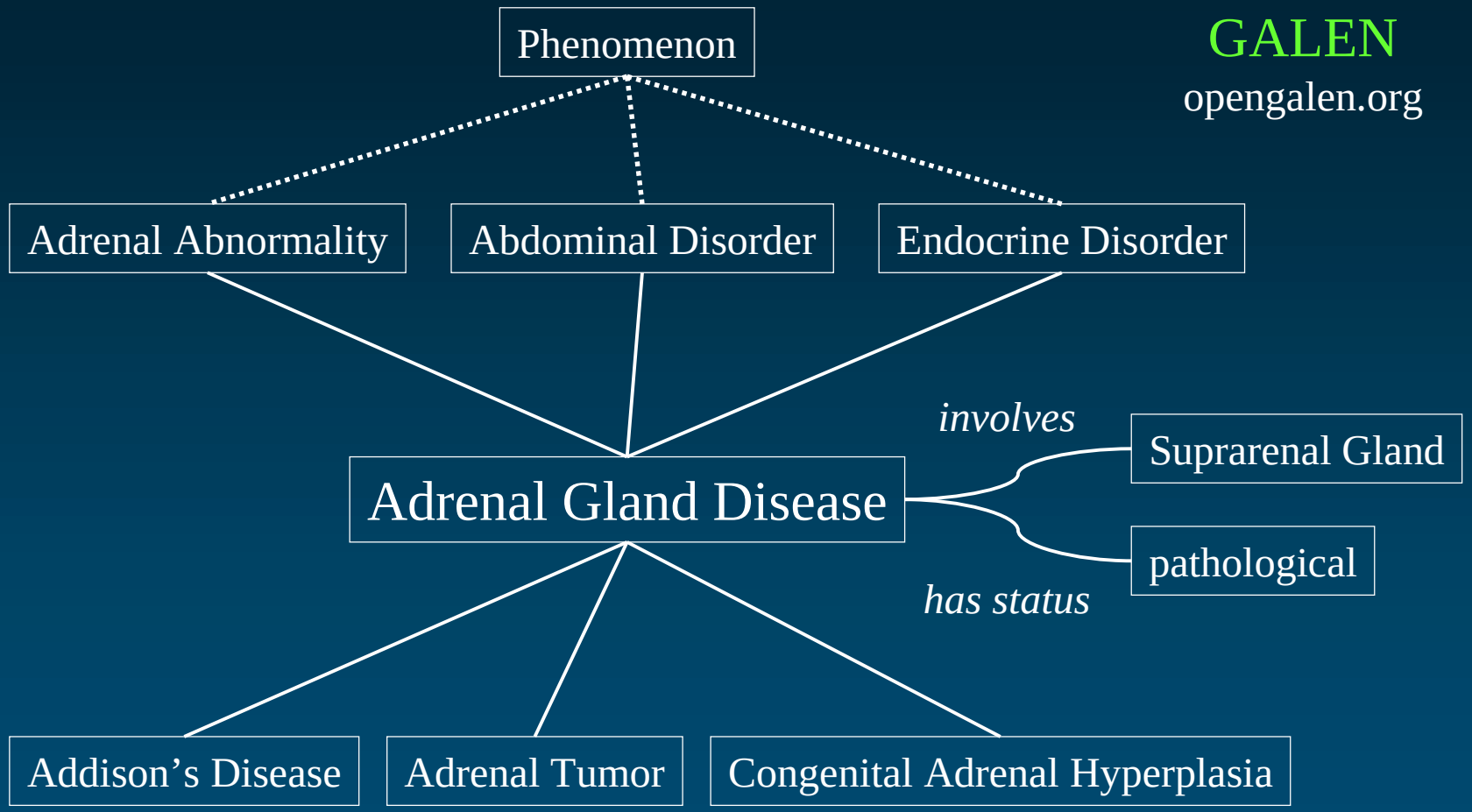
Medical Ontology Research

Terminology Adrenal gland diseases

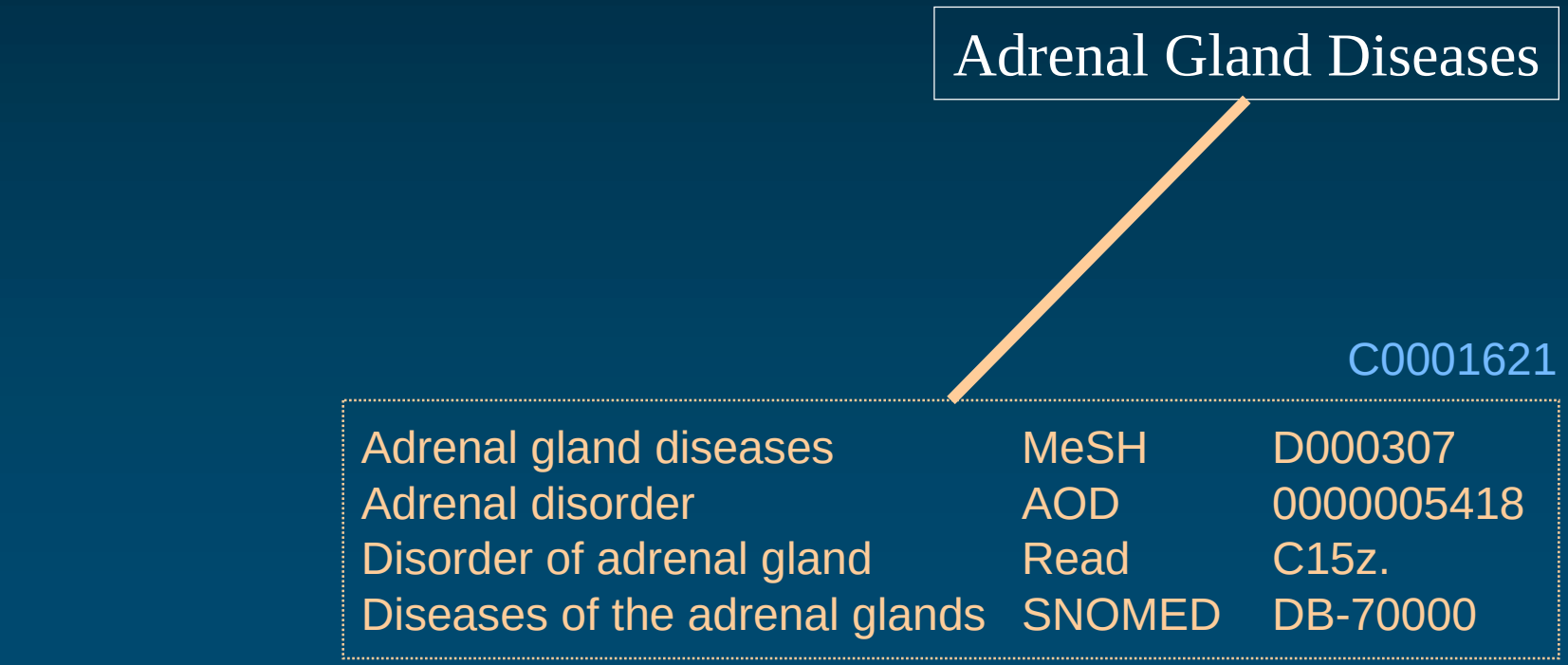


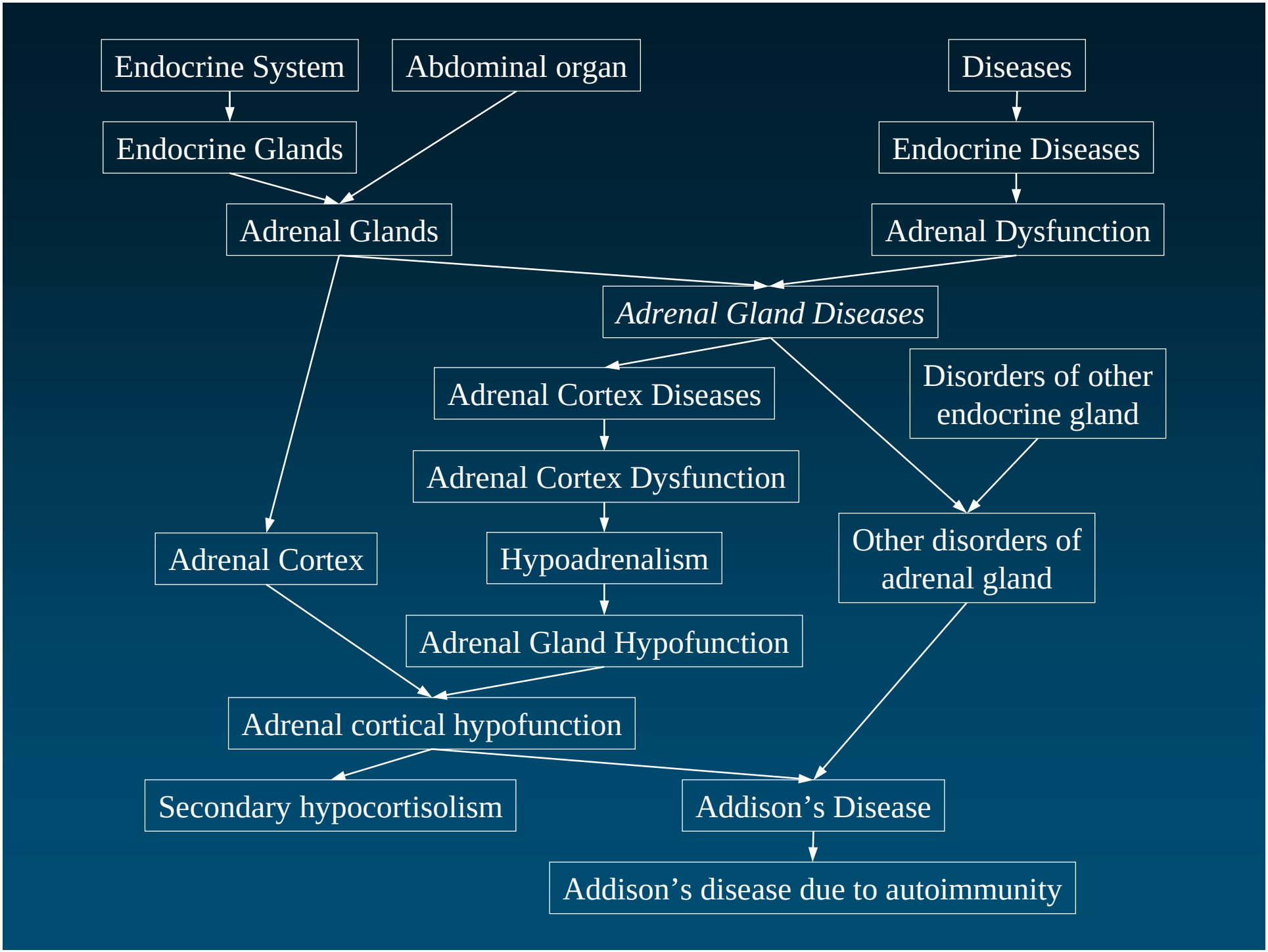
Adrenal gland diseases	MeSH	D000307
Adrenal disorder	AOD	0000005418
Disorder of adrenal gland	Read	C15z.
Diseases of the adrenal glands	SNOMED	DB-70000

Ontology Adrenal gland diseases

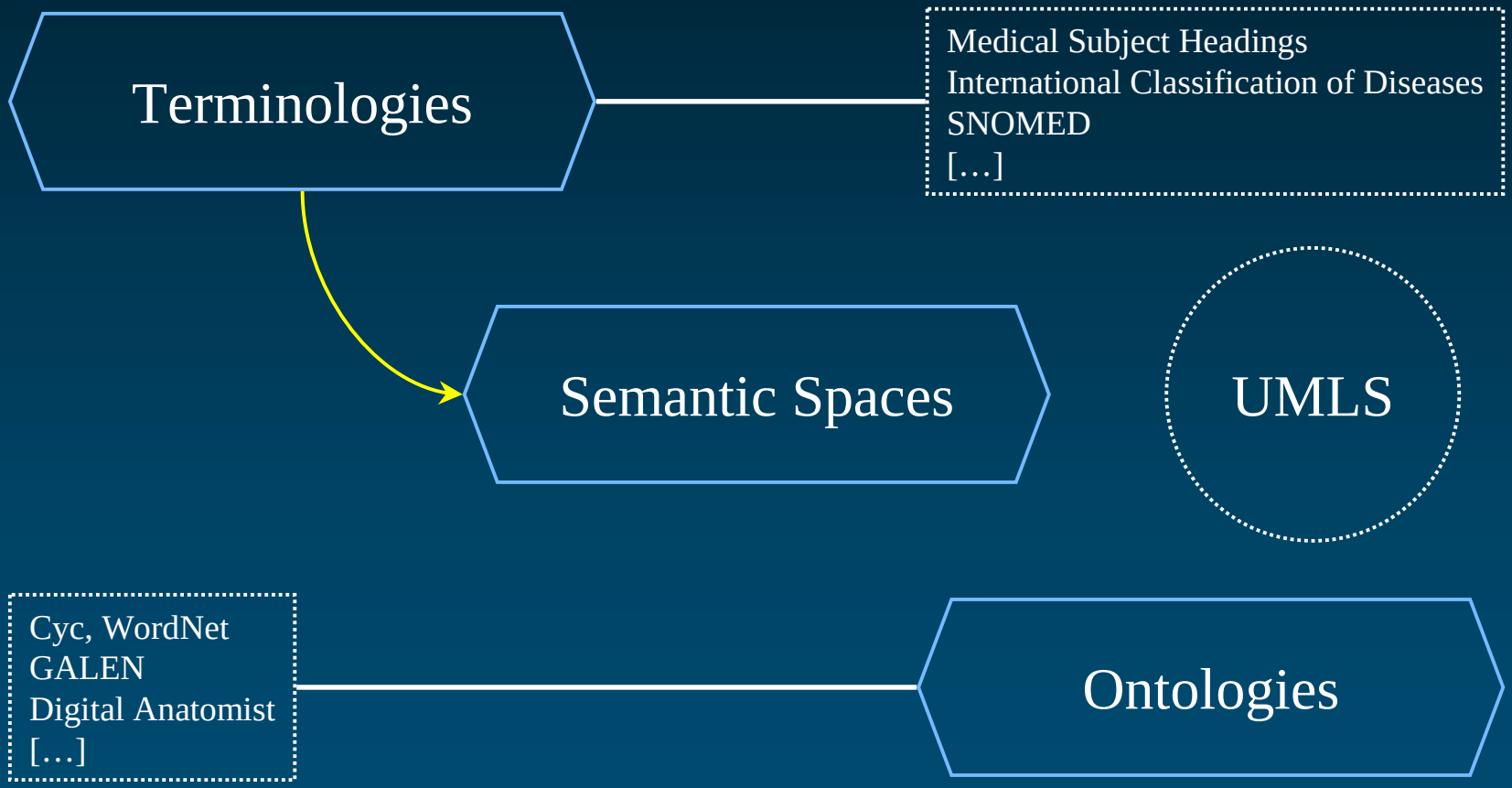


UMLS Adrenal gland diseases concept




Endocrine System
Abdominal organ
Diseases
Endocrine Glands
Endocrine Diseases
Adrenal Glands
Adrenal Dysfunction
Adrenal Gland Diseases
Adrenal Cortex Diseases
Disorders of other endocrine gland
Adrenal Cortex Dysfunction
Adrenal Cortex
Hypoadrenalism
Other disorders of adrenal gland
Adrenal Gland Hypofunction
Adrenal cortical hypofunction
Secondary hypocortisolism
Addison’s Disease
Addison’s disease due to autoimmunity

Biomedical knowledge organization



Outline

- Overview
- Benefits
- Limitations

Overview

Biomedical terminologies

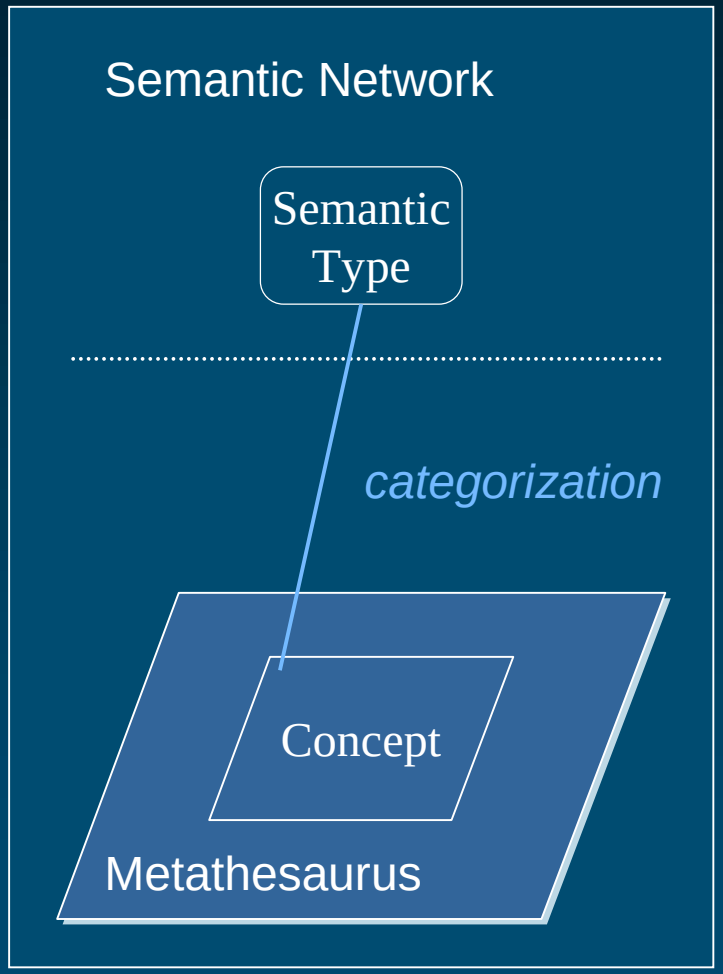
- Core vocabularies
  - anatomy (UWDA, Neuronames)
  - drugs (First DataBank, Micromedex)
  - medical devices (UMD, SPN)
- Several perspectives
  - clinical terms (SNOMED, CTV3)
  - information sciences (MeSH, CRISP)
  - administrative terminologies (ICD-9-CM, CPT-4)
  - standards (HL7, LOINC)

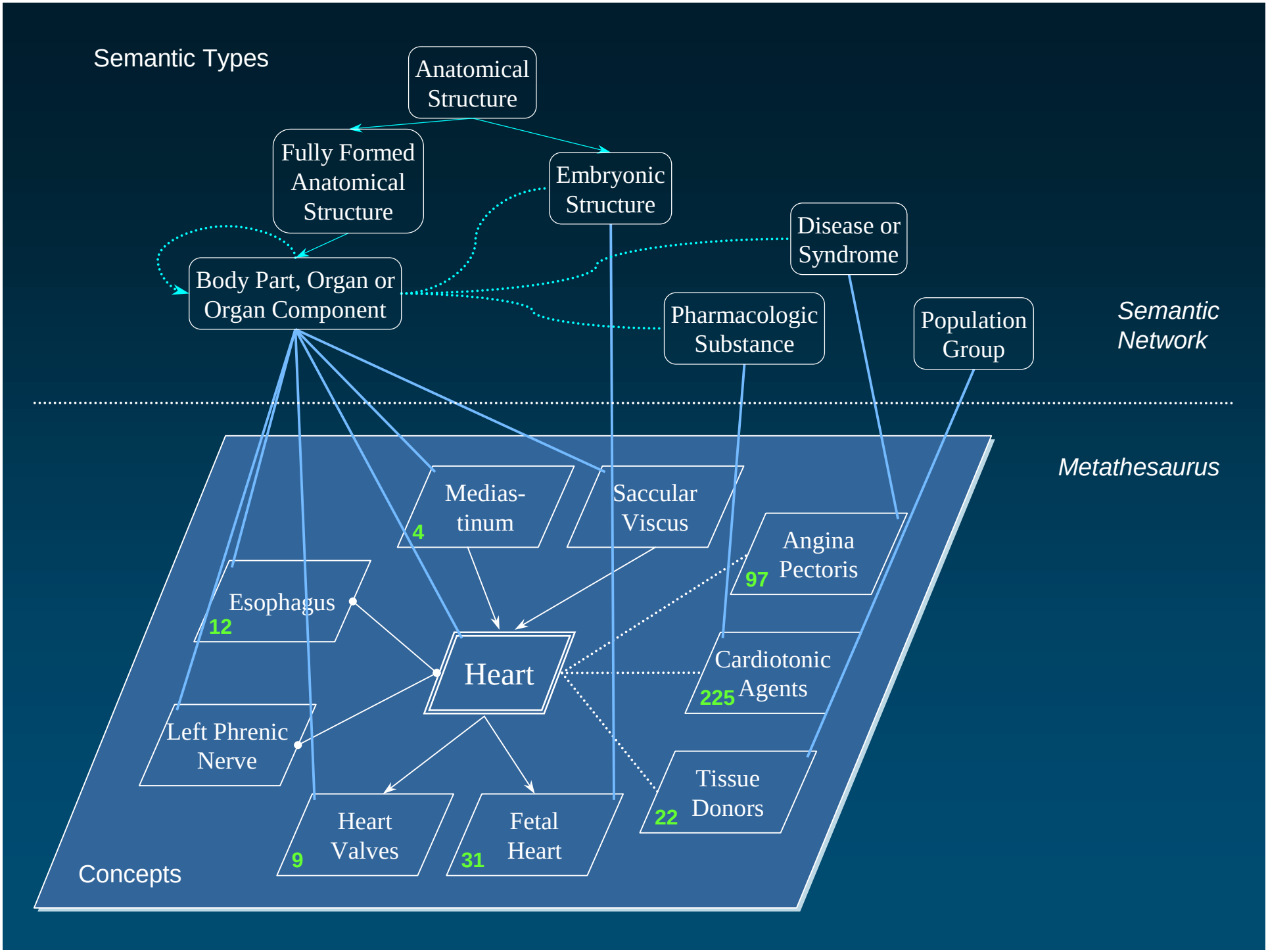
Biomedical terminologies (cont'd)

- Specialized vocabularies
  - nursing (NIC, NOC, NANDA, Omaha, PCDS)
  - dentistry (CDT)
  - oncology (PDQ)
  - psychiatry (DSM, APA)
  - adverse reactions (COSTART, WHO ART)
  - primary care (ICPC)
- Knowledge bases (AI/Rheum, DXplain, QMR)

UMLS

- Two-level structure
  - Semantic Network
    - 135 Semantic Types (STs)
    - 54 types of relationships among STs
  - Metathesaurus
    - 875,000 concepts
    - ~12 M inter-concept relationships
  - Link = categorization
- Lexical resources



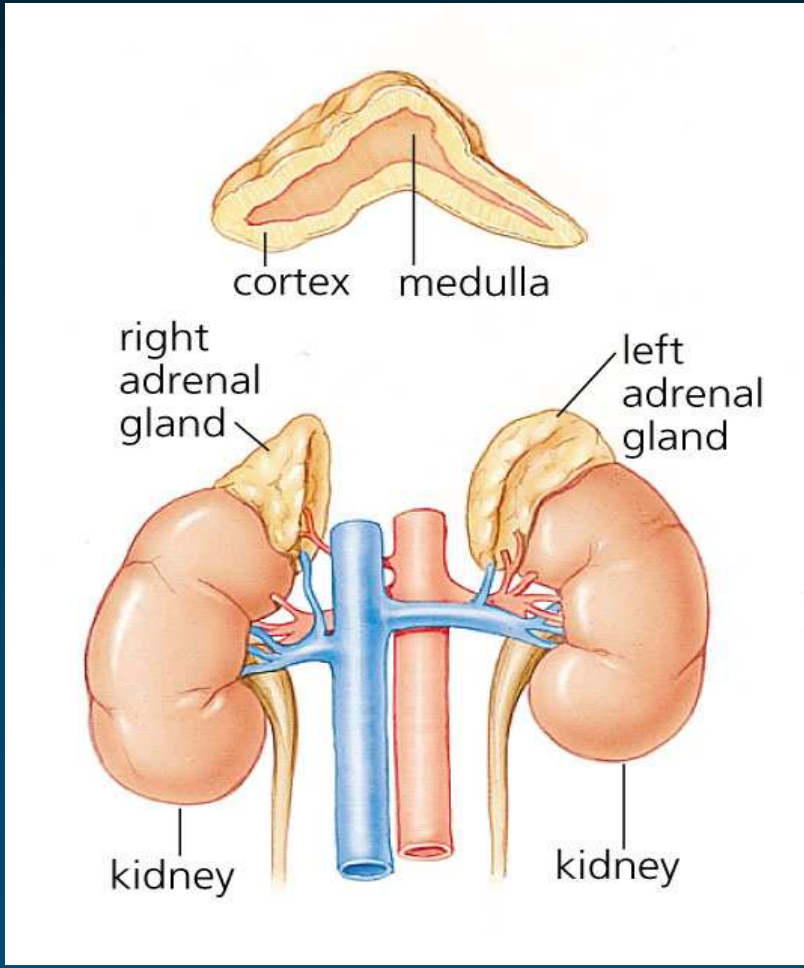

Semantic Types
Anatomical Structure
Fully Formed Anatomical Structure
Embryonic Structure
Disease or Syndrome
Body Part, Organ or Organ Component
Pharmacologic Substance
Population Group
Semantic Network
Metathesaurus
Medias-tinum
4
Saccular Viscus
Angina Pectoris
97
Esophagus
12
Heart
Cardiotonic Agents
225
Left Phrenic Nerve
Tissue Donors
22
Heart Valves
9
Fetal Heart
31
Concepts

UMLS Services

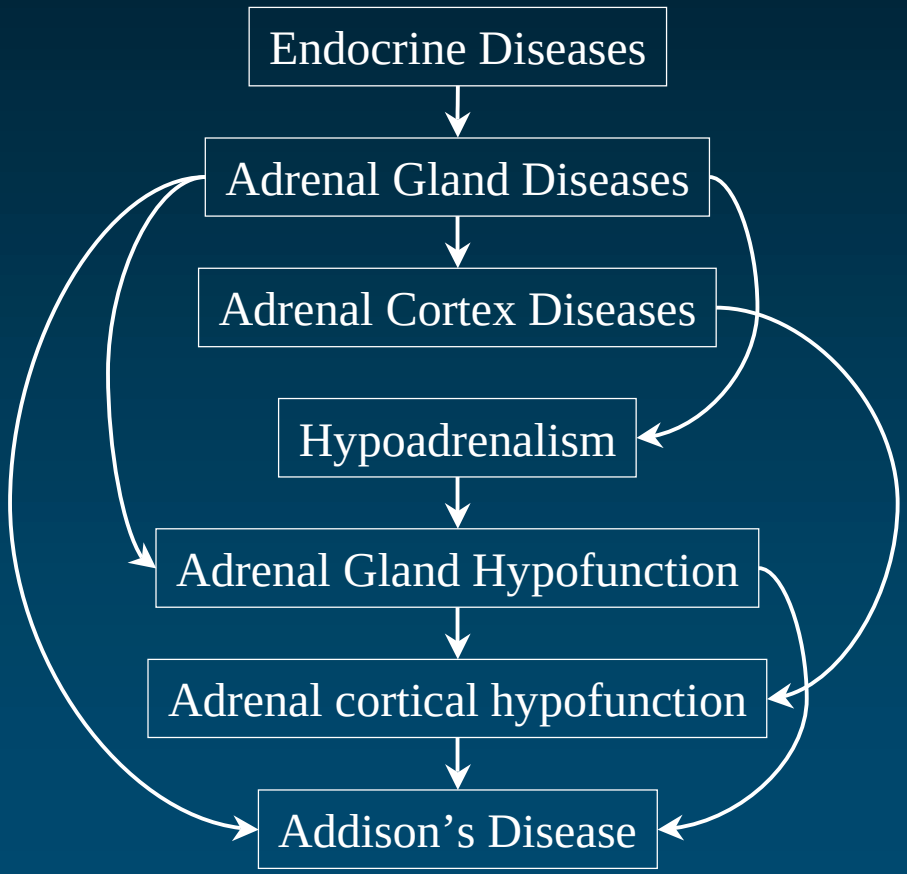
- Lexical tools (e.g., normalization)
- Browsers
- MetaMap
- API (Java, XML)

Addison’s disease

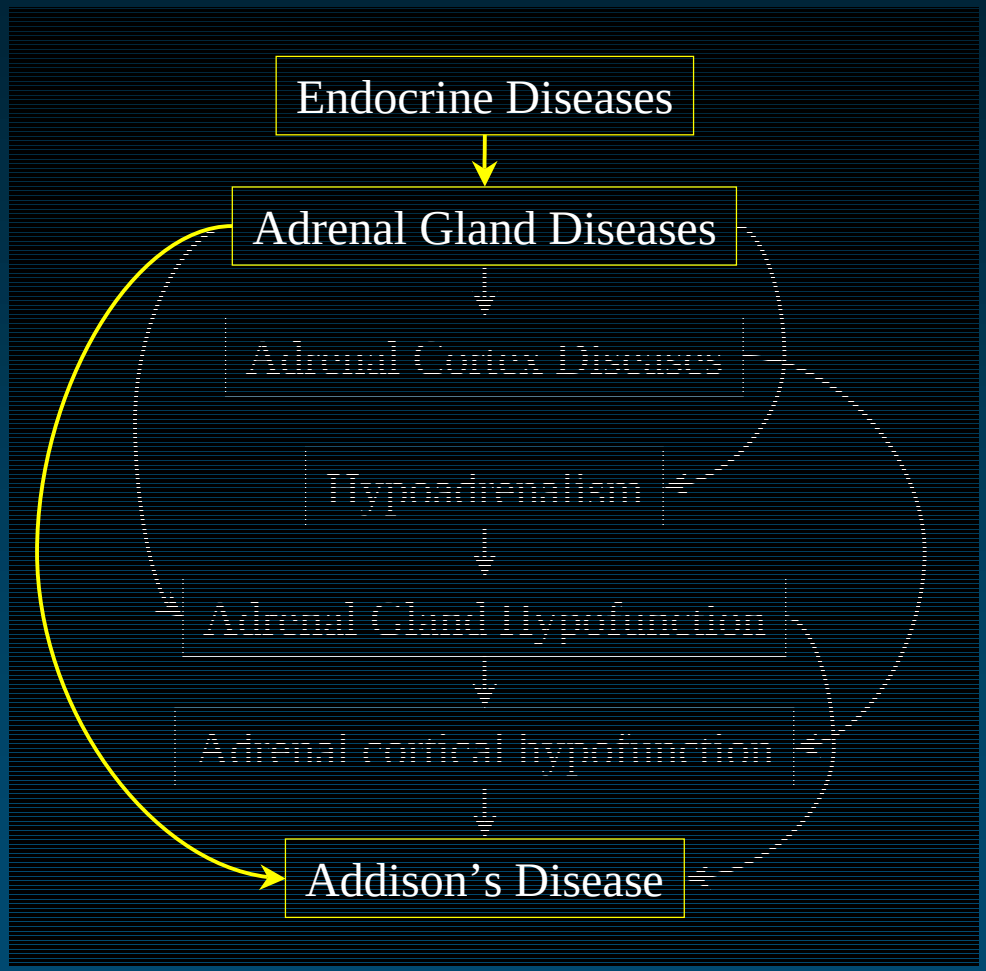
- Addison's disease is a rare endocrine disorder
- Addison's disease occurs when the adrenal glands do not produce enough of the hormone cortisol
- For this reason, the disease is sometimes called chronic adrenal insufficiency, or hypocortisolism



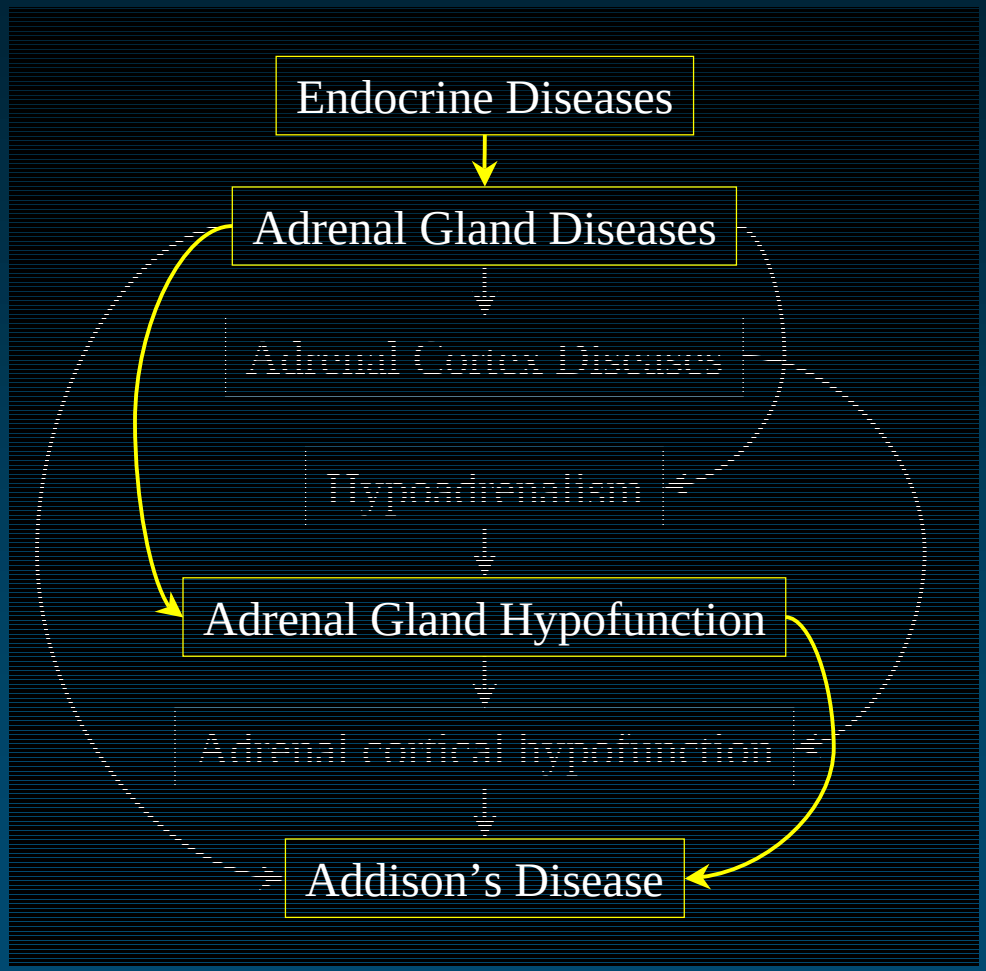
AD in UMLS Contexts



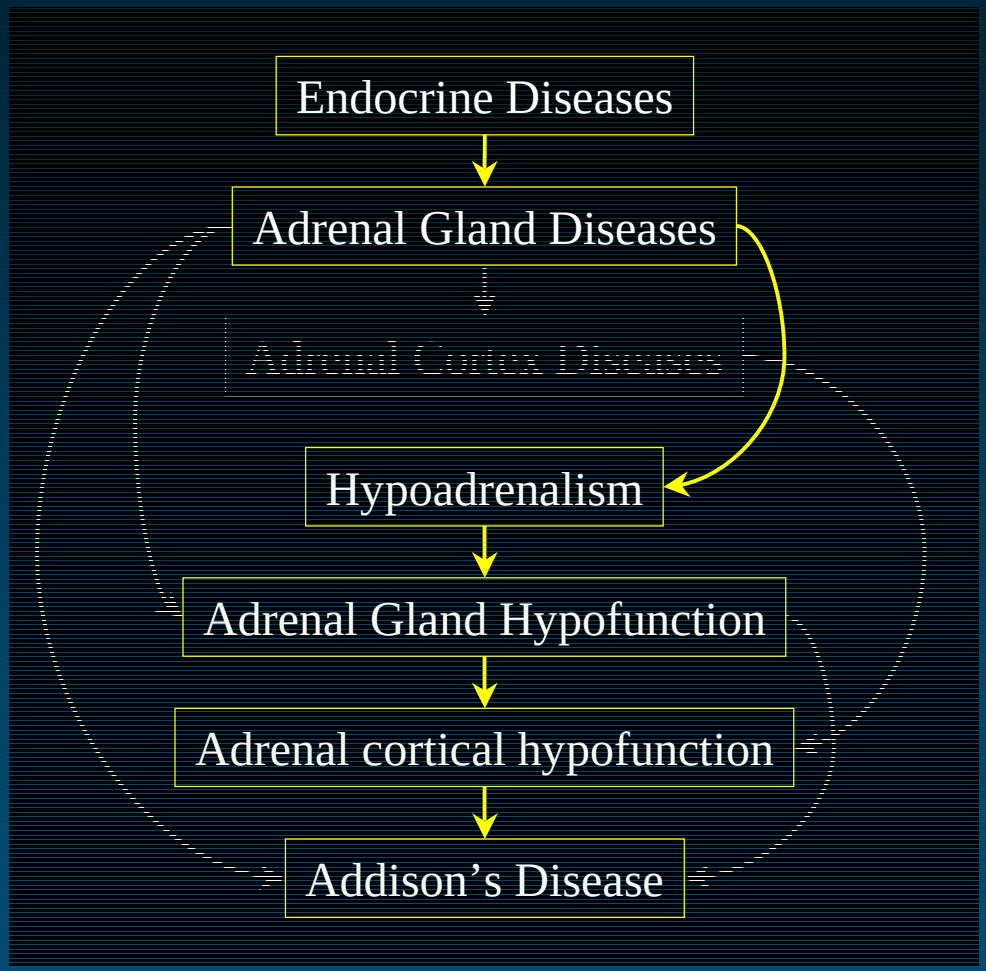
AD in UMLS SNOMED context



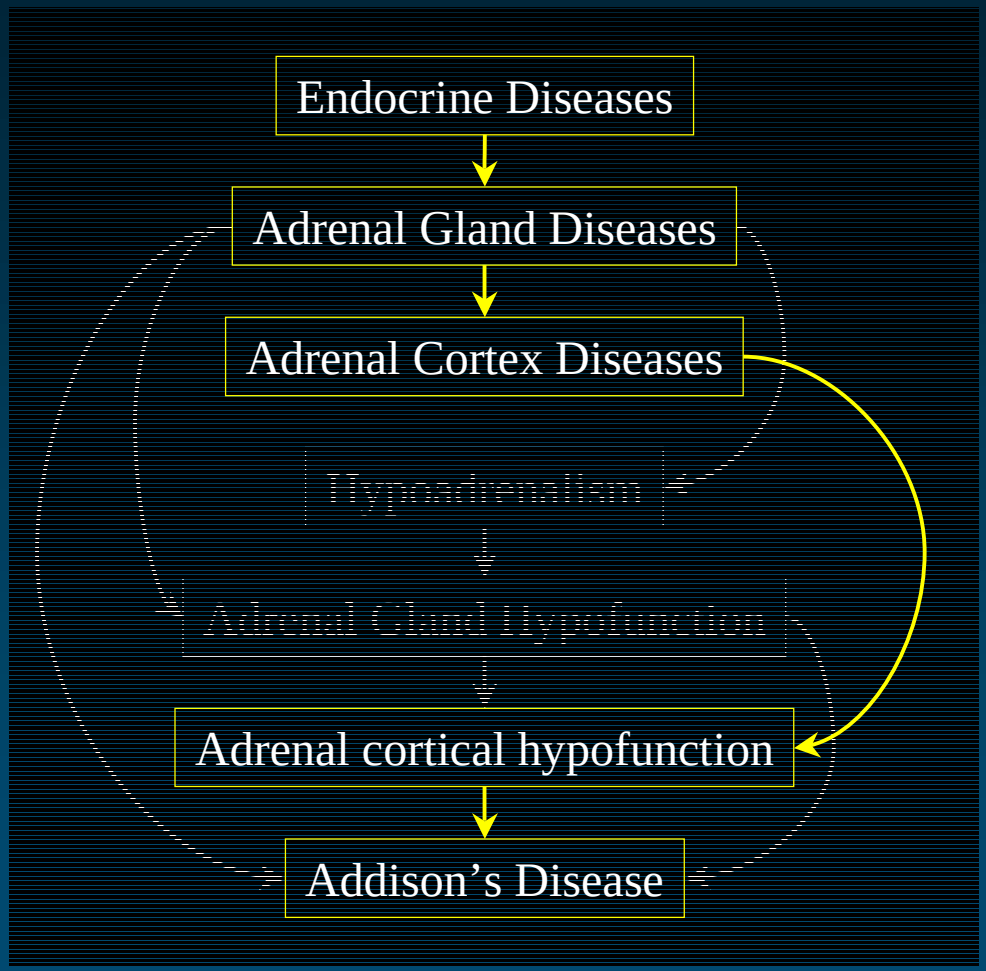
AD in UMLS MeSH context

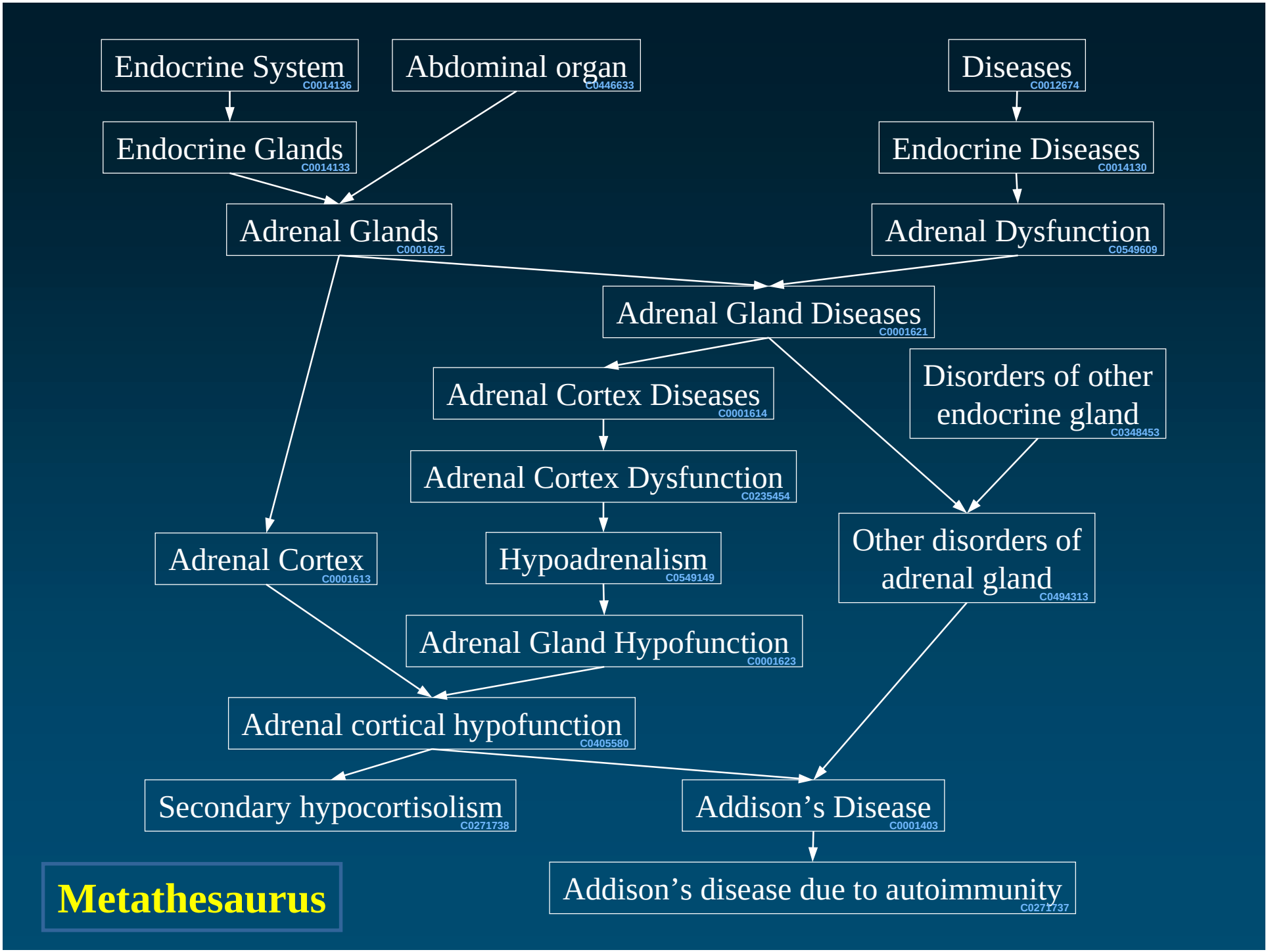


AD in UMLS Read Codes context



AD in UMLS AOD Thes. context




Endocrine System
C0014136
Abdominal organ
C0446633
Diseases
C0012674
Endocrine Glands
C0014133
Endocrine Diseases
C0014130
Adrenal Glands
C0001625
Adrenal Dysfunction
C0549609
Adrenal Gland Diseases
C0001621
Adrenal Cortex Diseases
C0001614
Disorders of other endocrine gland
C0348453
Adrenal Cortex Dysfunction
C0235454
Adrenal Cortex
C0001613
Hypoadrenalism
C0549149
Other disorders of adrenal gland
C0494313
Adrenal Gland Hypofunction
C0001623
Adrenal cortical hypofunction
C0405580
Secondary hypocortisolism
C0271738
Addison’s Disease
C0001403
Addison’s disease due to autoimmunity
C0271737
Metathesaurus

Benefits

UMLS compared to individual vocabularies

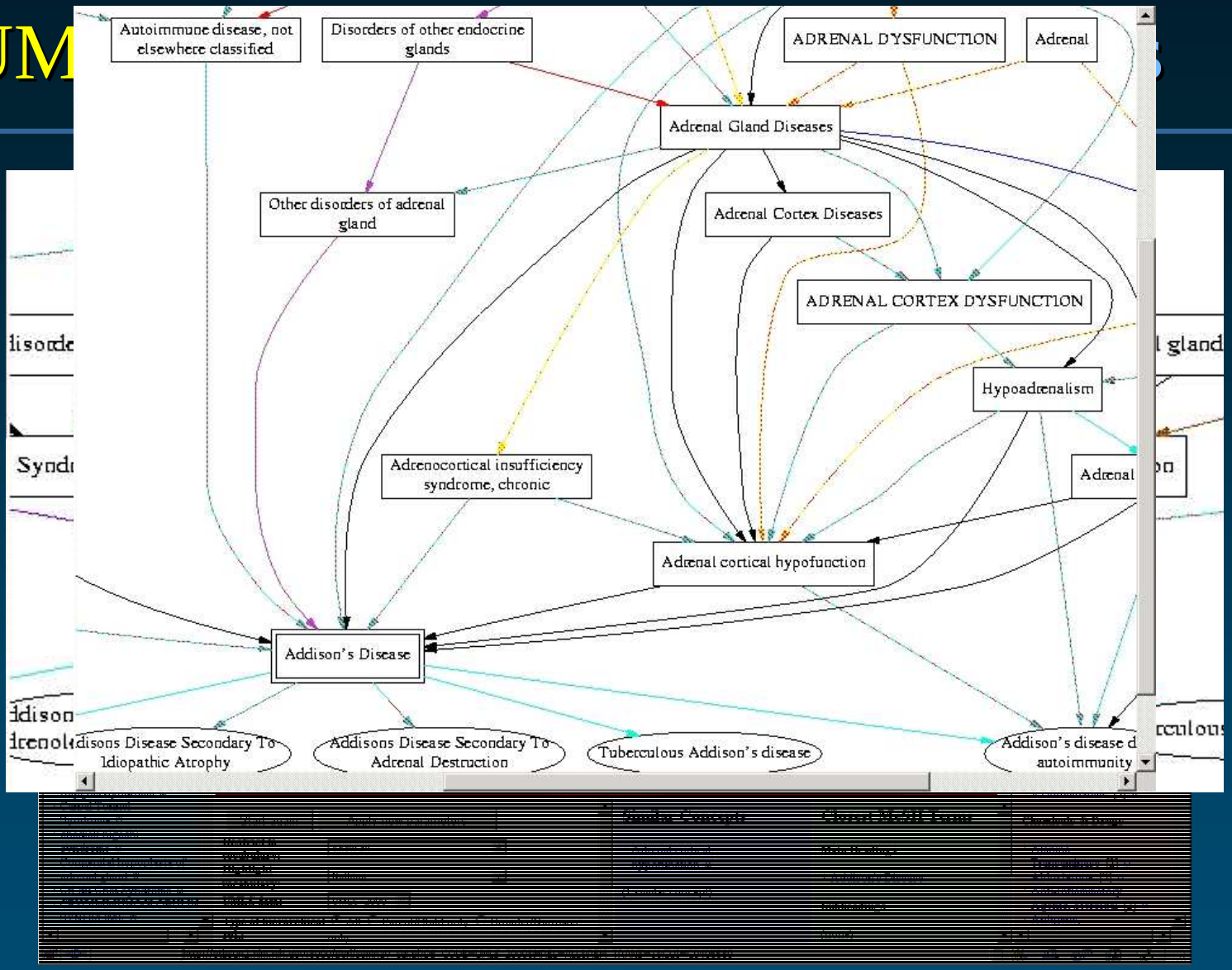
- Broader scope
- Extended coverage
- Finer granularity
- Unique identifier
- Synonymous terms clustered into concepts
- Additional synonyms
- Additional hierarchical relationships
- Semantic categorization

Direct benefits

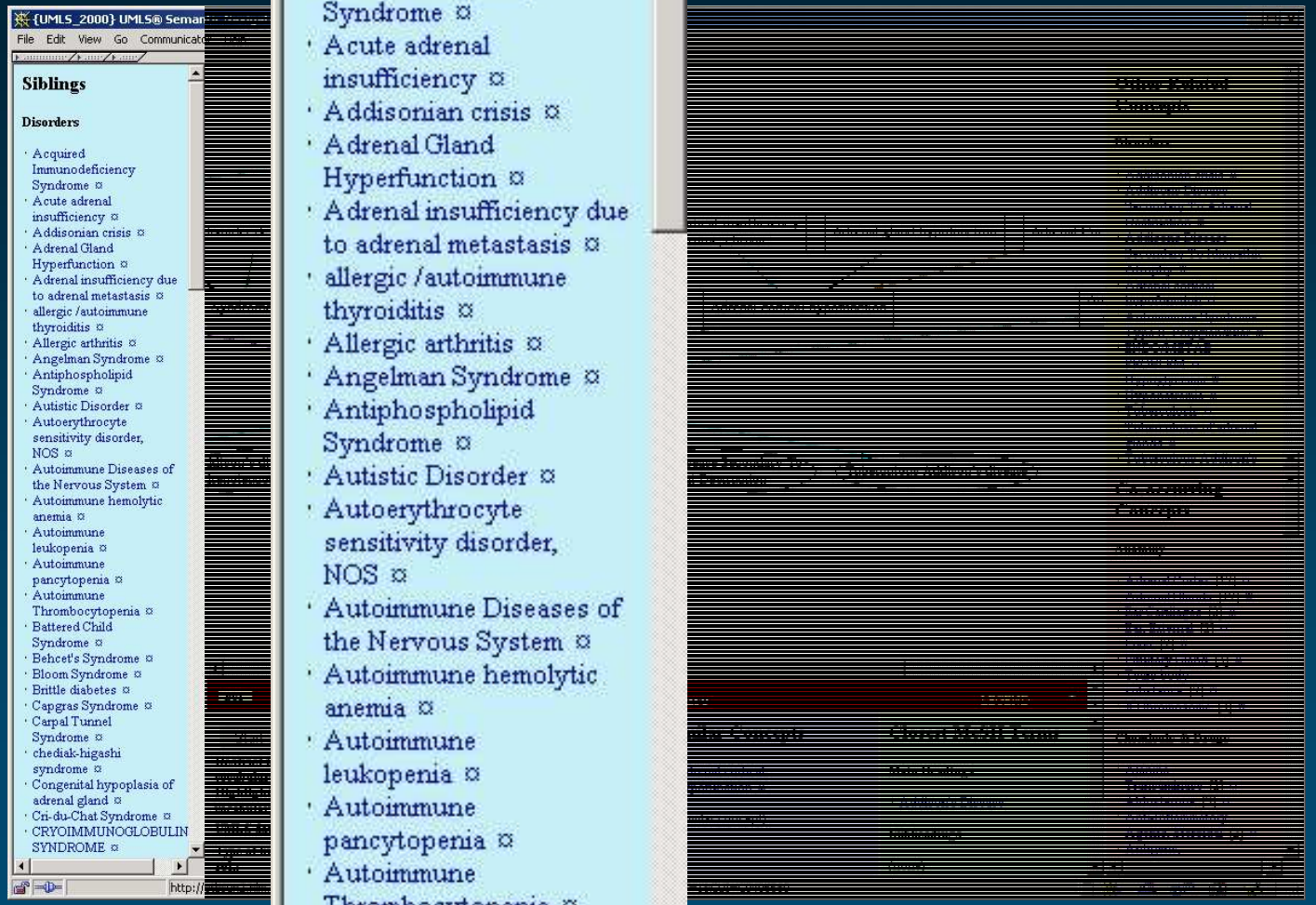
- Concept categorization
- Information retrieval
  - Synonyms
  - Cross-language features
- Information extraction
  - MetaMap
  - Normalization
- Information visualization
  - Knowledge Source Server
  - Semantic Navigator

UMLS Semantic Navigator

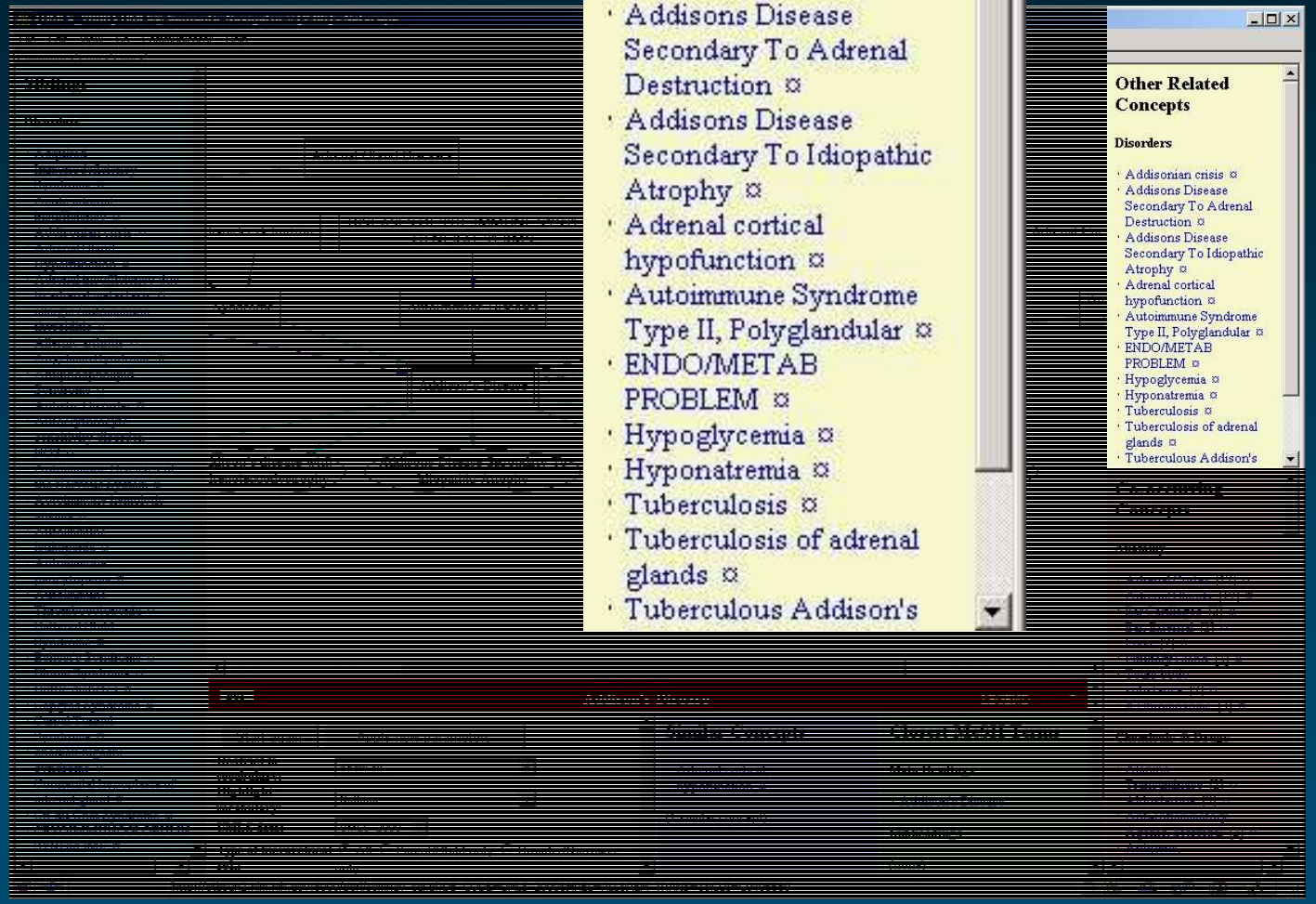
UM


Autoimmune disease, not elsewhere classified
Disorders of other endocrine glands
ADRENAL DYSFUNCTION
Adrenal
Adrenal Gland Diseases
Other disorders of adrenal gland
Adrenal Cortex Diseases
ADRENAL CORTEX DYSFUNCTION
Hypoadrenalism
Adrenocortical insufficiency syndrome, chronic
Adrenal
Adrenal cortical hypofunction
Addison's Disease
Addisons Disease Secondary To Idiopathic Atrophy
Addisons Disease Secondary To Adrenal Destruction
Tuberculous Addison's disease
Addison's disease d autoimmunity

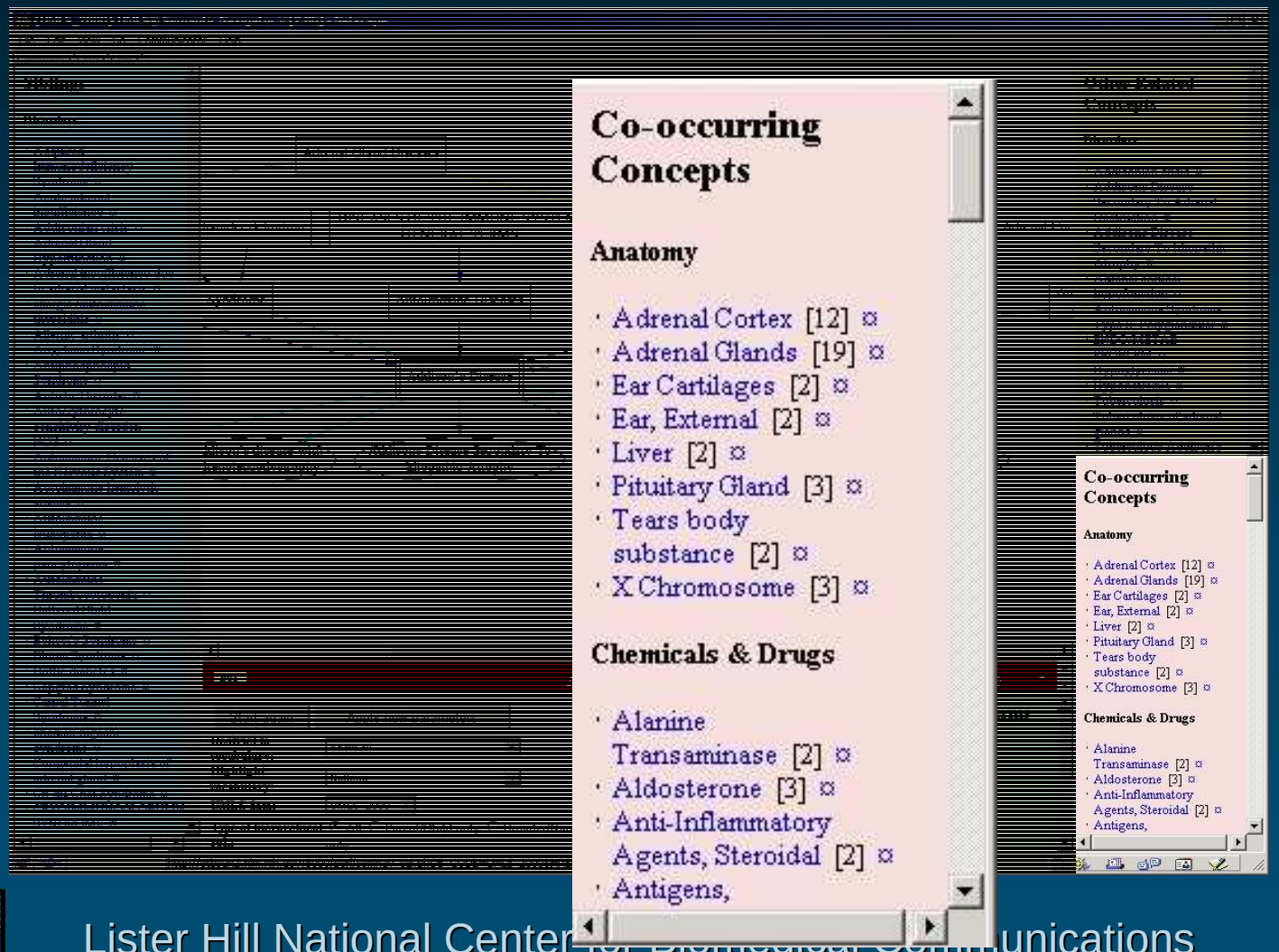
UMLS Se igator Concepts



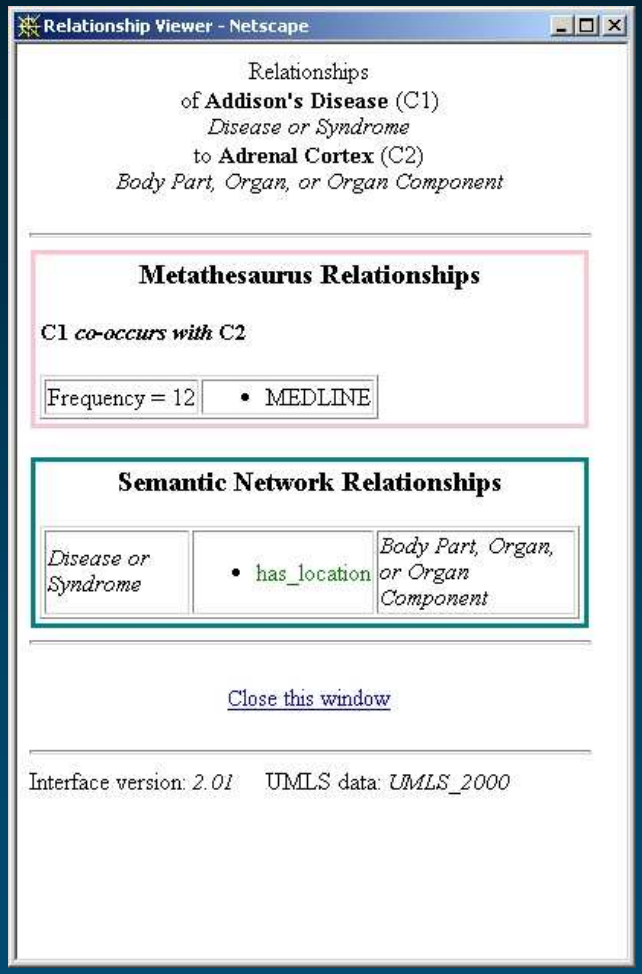
UMLS Semantic N oncepts

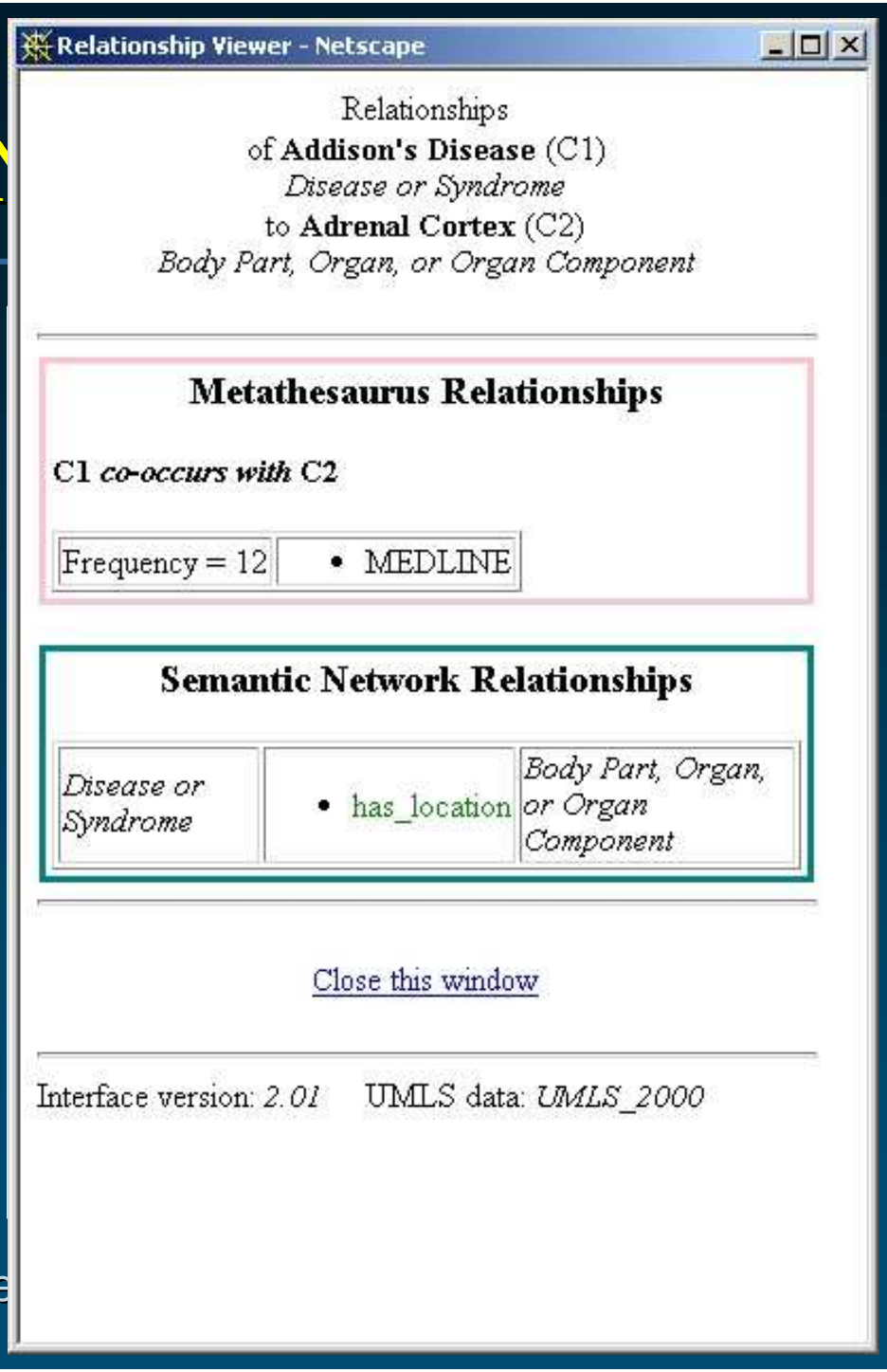


UMLS Semantic Navigator Concepts



UMLS Semantic N





NATIONAL LIBRARY OF MEDICINE

Lister Hill National Cente

Additional (indirect) benefits

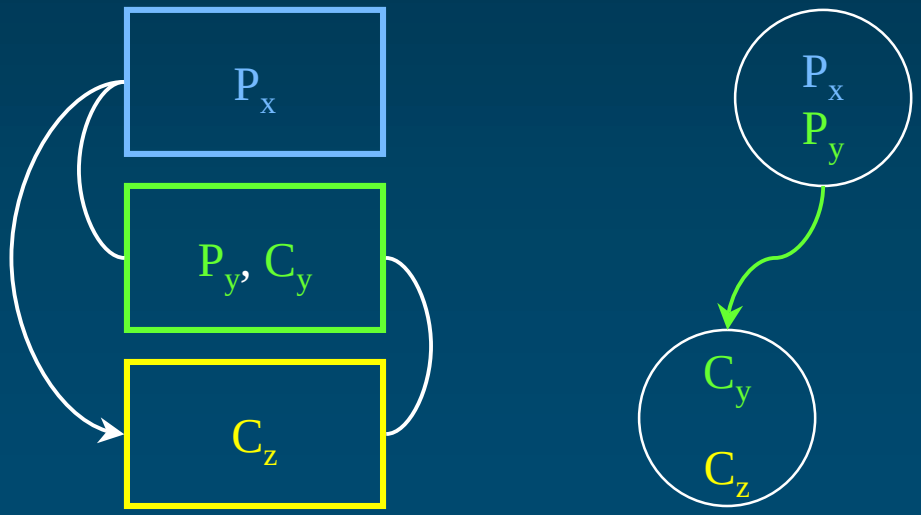
- Examples
  - Mapping across vocabularies
  - Semantics of statistical associations
  - Redundancy in hierarchical relations

Additional (indirect) benefits

Mapping across vocabularies

Terminology integration

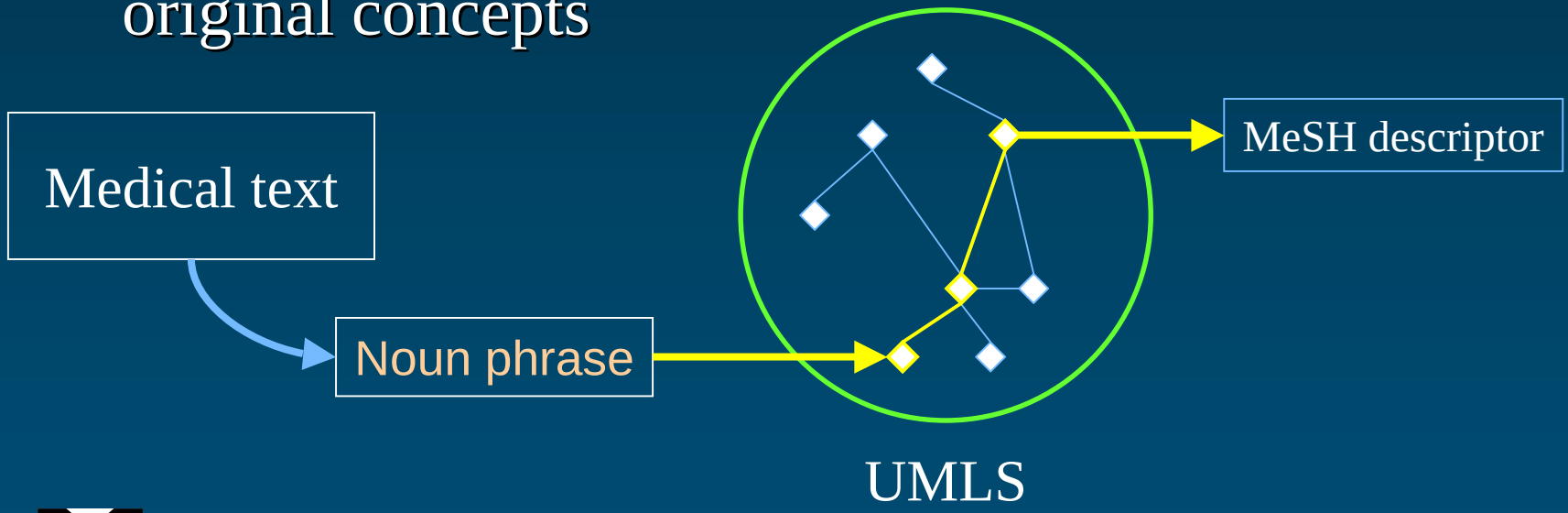
- Terminology integration is a step towards interoperability
  - Clusters of synonyms from different sources
  - Paths between terms from different sources



Indexing Initiative

[Aronson & al., *AMIA*, 2000]

- For noun phrases extracted from medical texts, map to UMLS concepts
- Then, select from the MeSH vocabulary the concepts that are the most closely related to the original concepts



Restrict to MeSH

[Bodenreider & al., *AMIA*, 1998]

- Based on the principle of semantic locality
- Use different components of the UMLS
- 4 techniques of increasing aggressiveness
  - Use Synonymy — MRCON + MRSO
  - Use Associated expressions (ATXs) — MRATX
  - Explore the Ancestors — MRREL + SN
  - Explore the Other related concepts — MRREL + SN

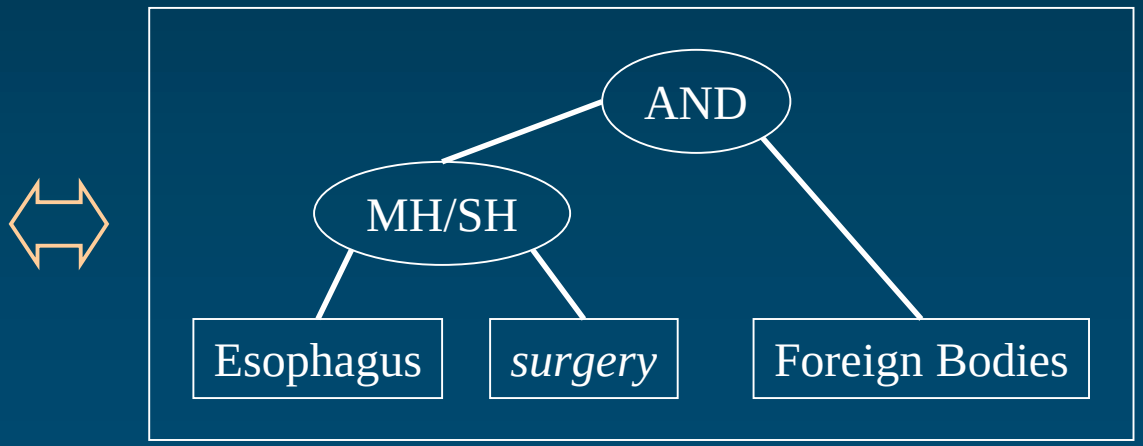
Restrict to MeSH: Synonymy

- Term mapped to Source concept
- For this concept, is there a synonym term that comes from MeSH? (MRSO)

Restrict to MeSH: Assoc. expressions

- If not,
- Is there an associated expression (ATX) that describes this concept using a combination of MeSH descriptors? (MRATX)

Endoscopic removal of intraluminal foreign body from oesophagus without incision



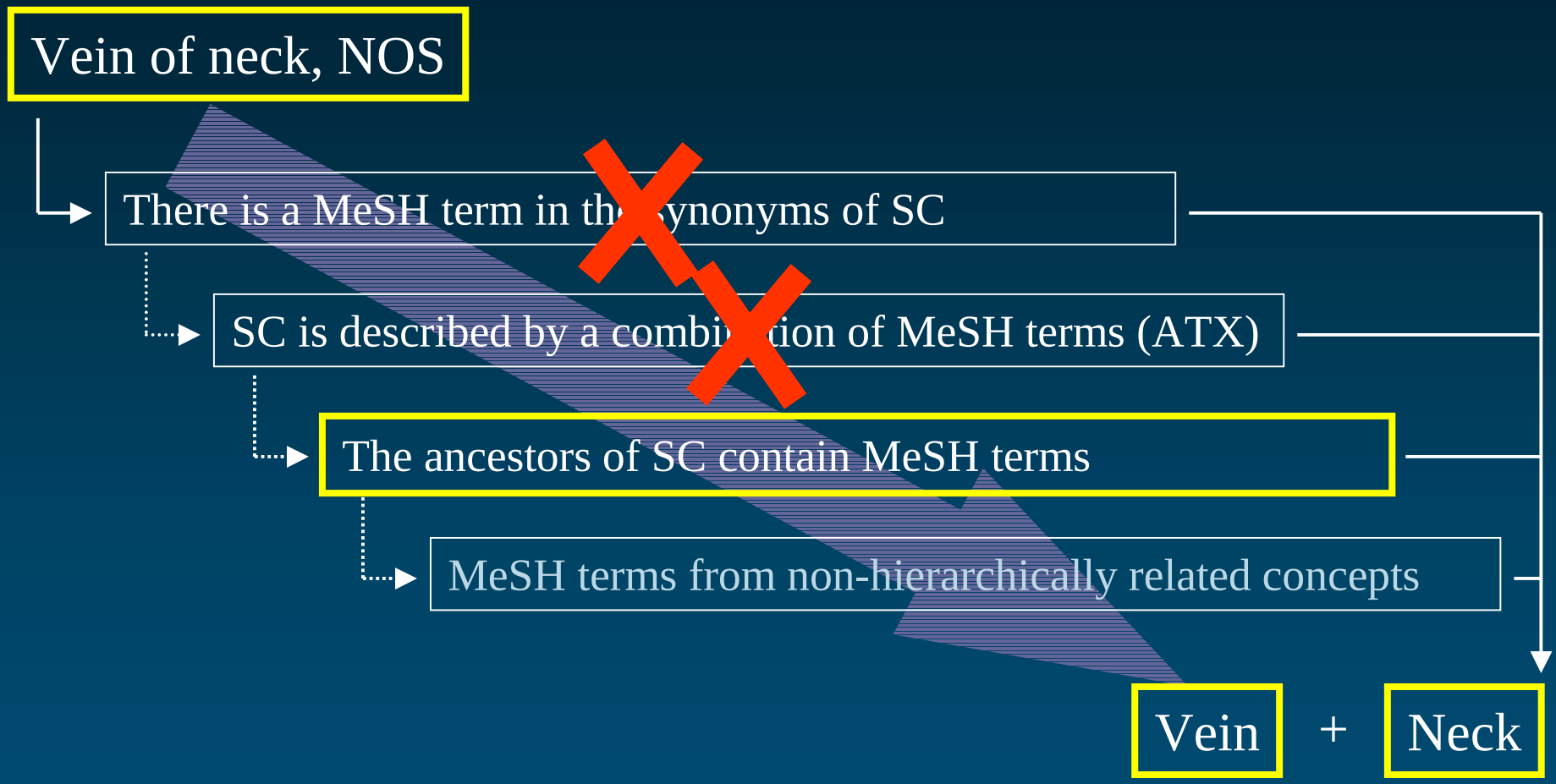
Restrict to MeSH: Ancestors

- If not, let us build the graph of the ancestors of this concept
  - using parents and broader concepts (MRREL)
  - all the way to the top
  - excluding ancestors whose semantic types are not compatible with those of the source concept (MRSTY)
- From the graph, select the concepts that come from MeSH (MRSO)
- Remove those that are ancestors of another concept coming from MeSH

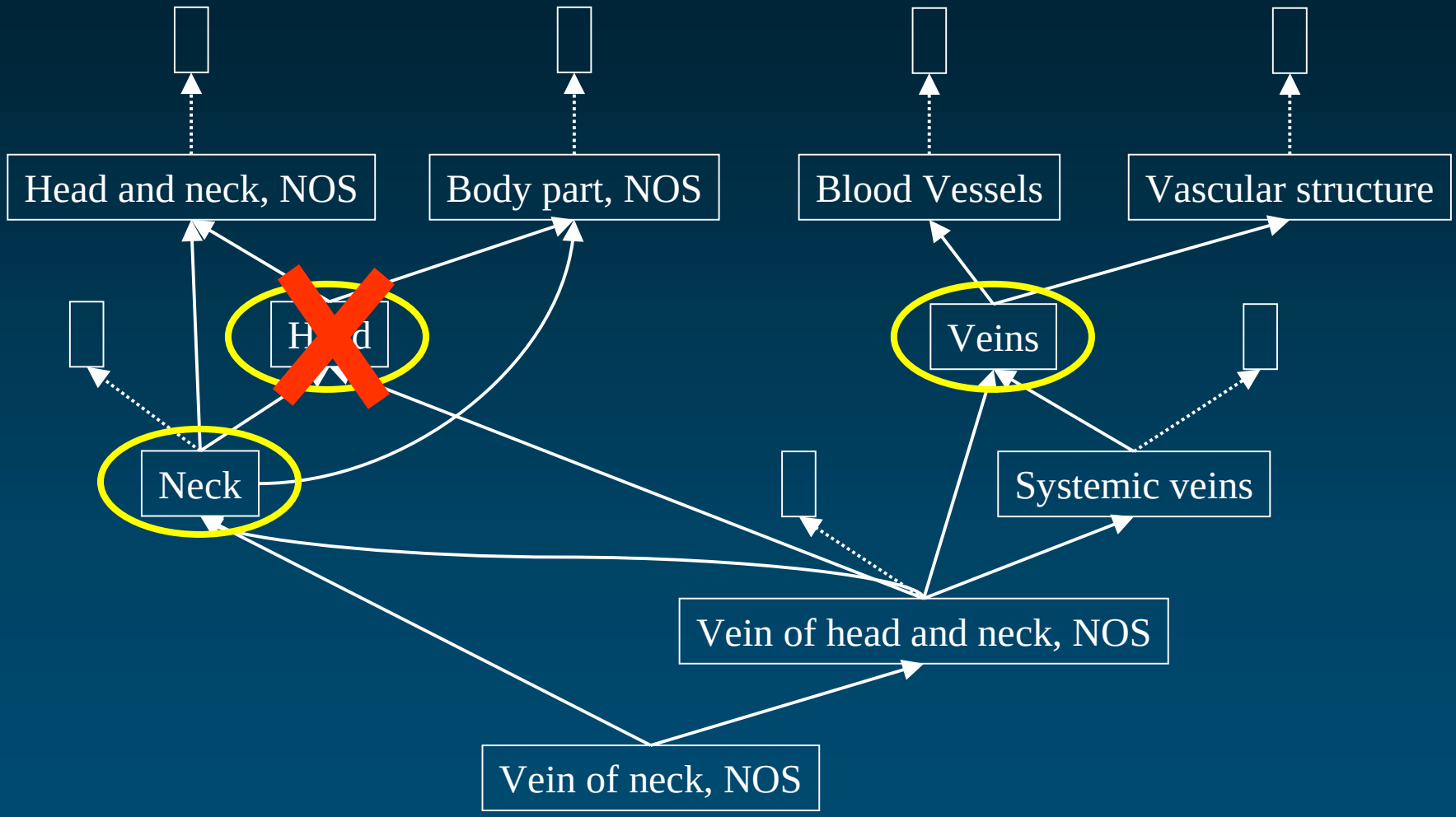
Restrict to MeSH: Other related concepts

- If not, explore the other related concepts (MRREL) whose semantic types are compatible with those of the source concept (MRSTY)
- From those, select the concepts that come from MeSH (MRSO)

Restrict to MeSH: Example



Restrict to MeSH: Example



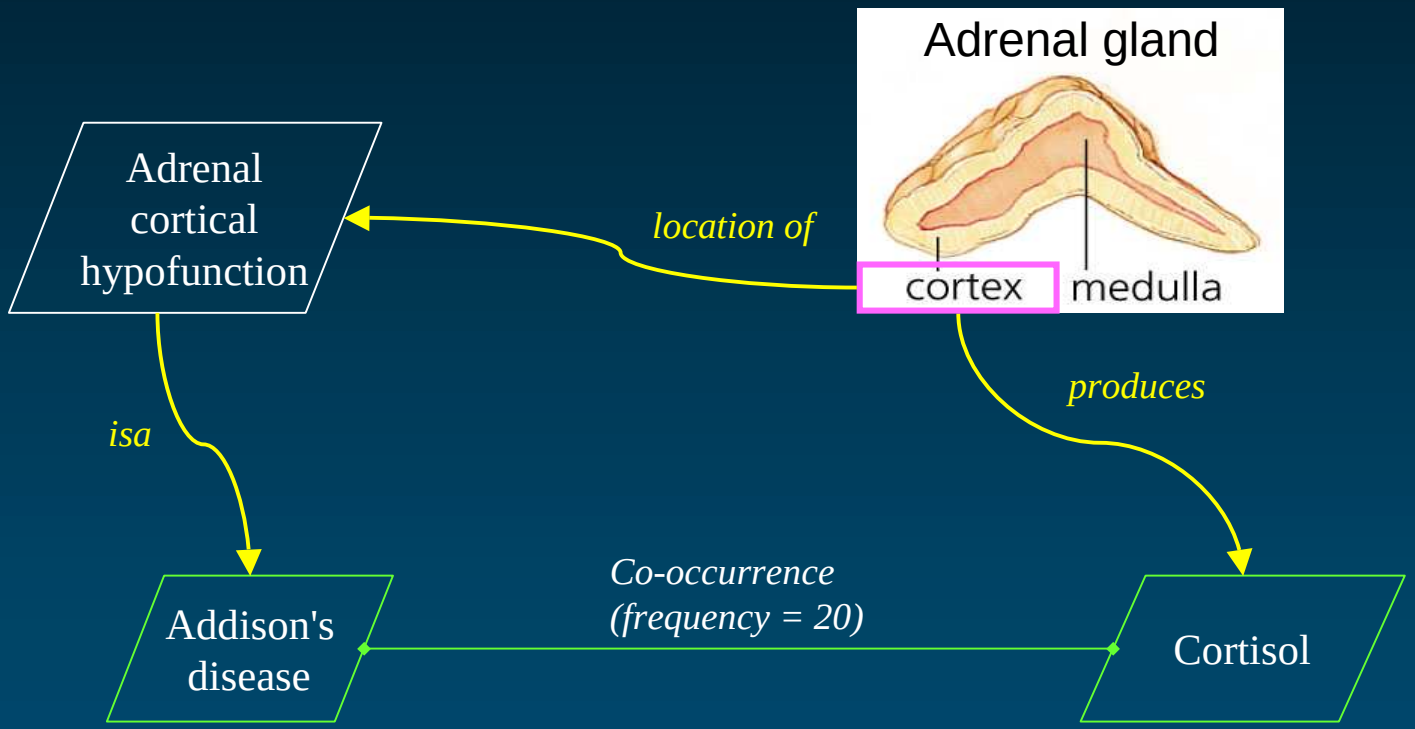
Additional (indirect) benefits

Semantics
of statistical associations

Co-occurrence Overview [Burgun & al., *MEDINFO*, 2001]

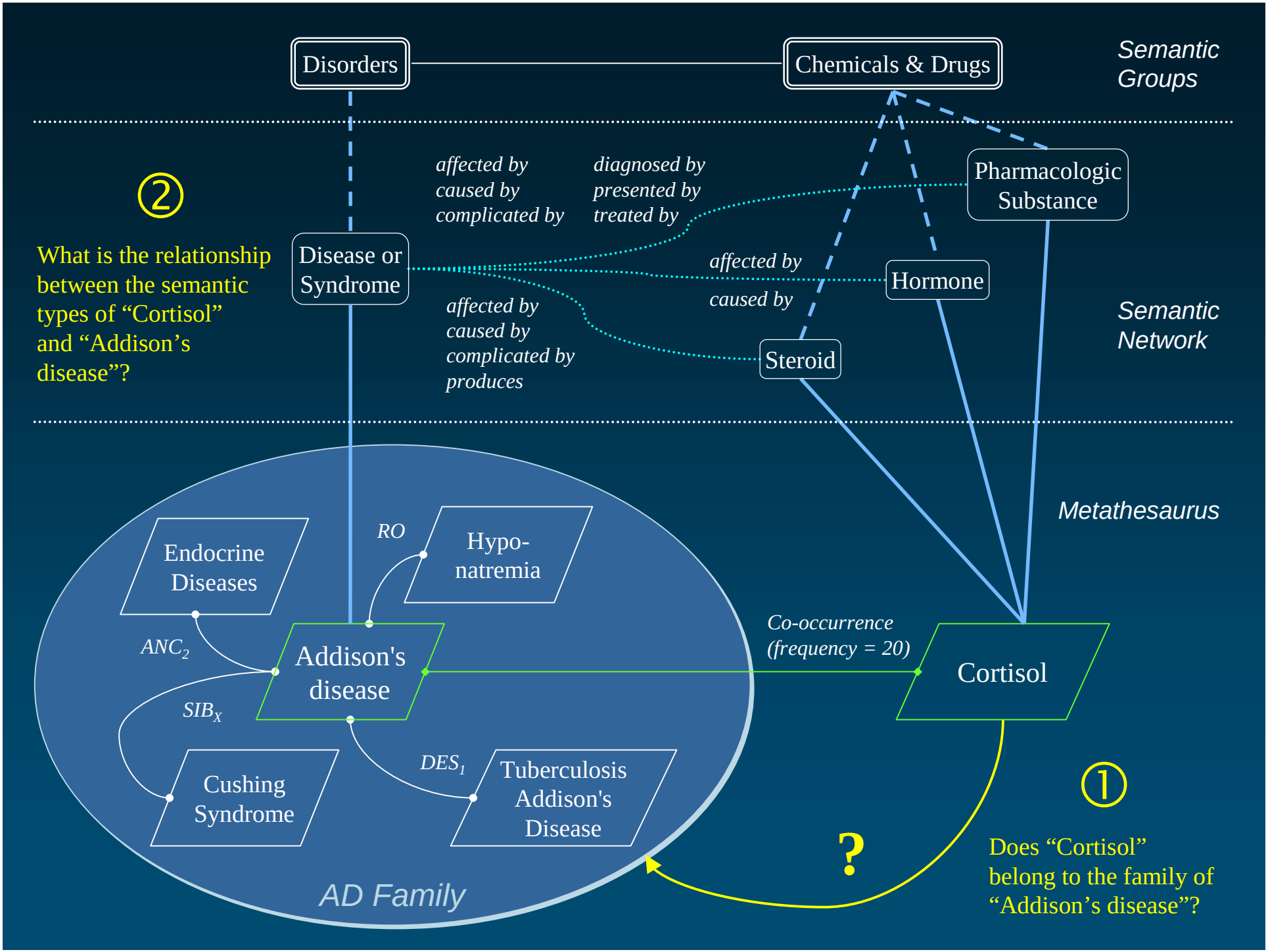
- Co-occurrence between MeSH descriptors in MEDLINE citations
- 8 M pairs of co-occurring concepts
- Implicit semantics
- The UMLS provides knowledge for helping make this relationship explicit
  - Corresponding symbolic knowledge (Metathesaurus)
  - Categorization (Semantic Network)

Co-occurrence Example



Co-occurrence Methods

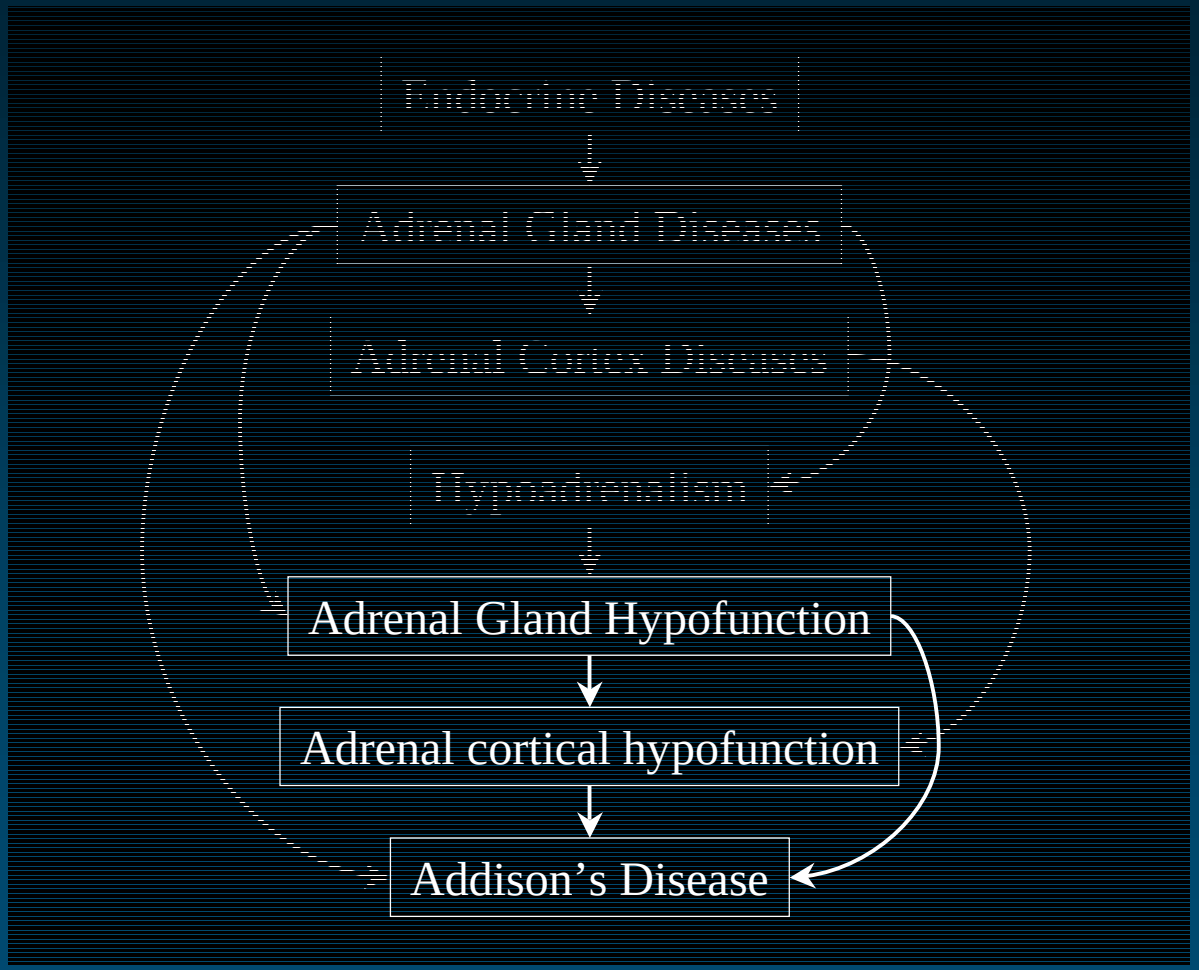
- Based on Metathesaurus relationships
  - Does “Cortisol” belong to the family of “Addison’s disease”?
- Based on Semantic Network relationships
  - What is the relationship between the semantic types of “Cortisol” and “Addison’s disease”?


Disorders
Chemicals & Drugs
Semantic Groups
affected by
caused by
complicated by
diagnosed by
presented by
treated by
Pharmacologic Substance
②
What is the relationship between the semantic types of “Cortisol” and “Addison’s disease”?
Disease or Syndrome
affected by
caused by
Hormone
affected by
caused by
complicated by
produces
Steroid
Semantic Network
Metathesaurus
Endocrine Diseases
RO
Hypo-natremia
ANC2
Addison's disease
Co-occurrence (frequency = 20)
Cortisol
SIBX
DES1
Cushing Syndrome
Tuberculosis Addison's Disease
AD Family
?
①
Does “Cortisol” belong to the family of “Addison’s disease”?

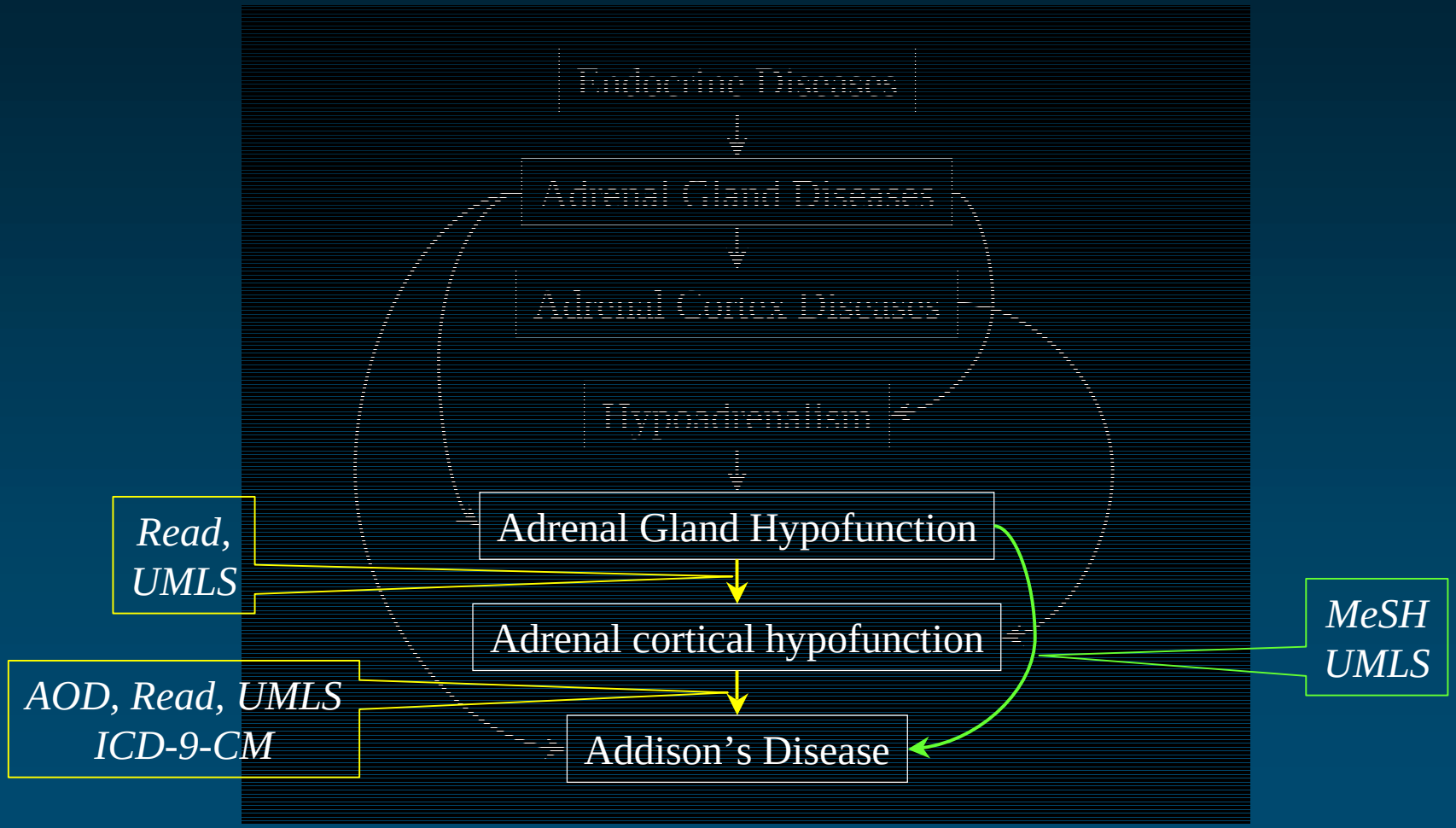
Additional (indirect) benefits

Redundancy
in hierarchical relations

AD in UMLS Redundancy

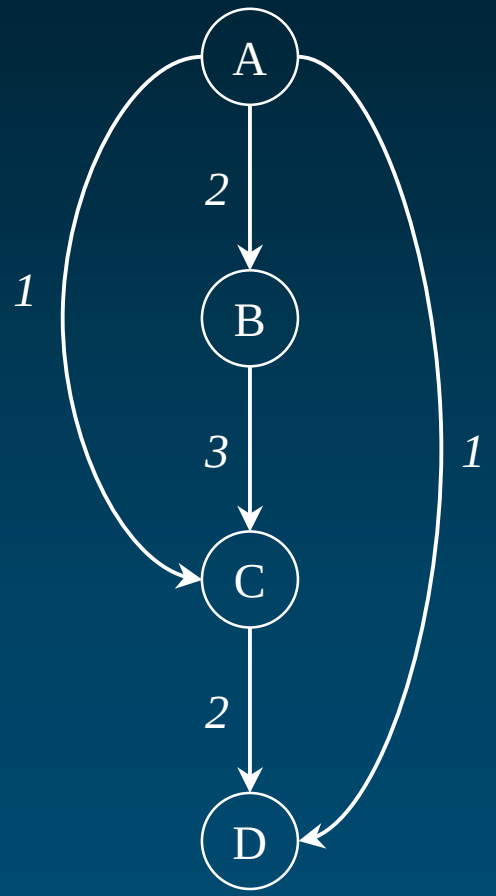


AD in UMLS Redundancy



Redundancy index

[Bodenreider & al., *AMIA*, 2003 (subm.)]



$IR_{AD} = IRP_{AD} + IRP_{ACD} + IRP_{ABCD} = 4$

with:

$IRP_{AD} = NS_{AD} = 1$

$IRP_{ACD} = \min(NS_{AC}, NS_{CD}) = \min(1, 2) = 1$

$IRP_{ABCD} = \min(NS_{AB}, NS_{BC}, NS_{CD}) = \min(2, 3, 2) = 2$

Limitations

Limitations

[Cimino, *JAMIA*, 1998]

- Structural inconsistency
  - Cycles in the graph of hierarchical relations
- Semantic inconsistency
  - Between Metathesaurus and Semantic Network
  - Meaning of *isa*
- Missing relations
  - Synonymy
  - Hierarchical relations (missing or underspecified)

Compensation mechanisms

- Examples
  - Removing cycles from hierarchical relations
  - Lexically-suggested hyponymic relations

Compensation mechanisms

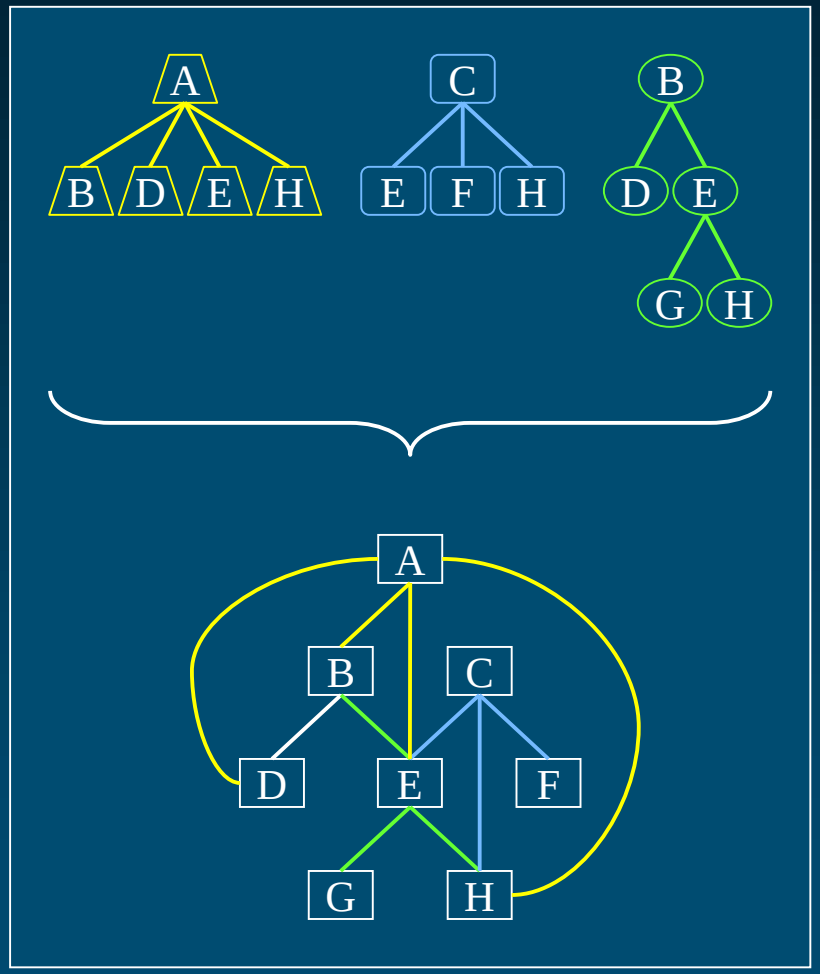
Removing cycles
from hierarchical relations

Hierarchies in source vocabularies

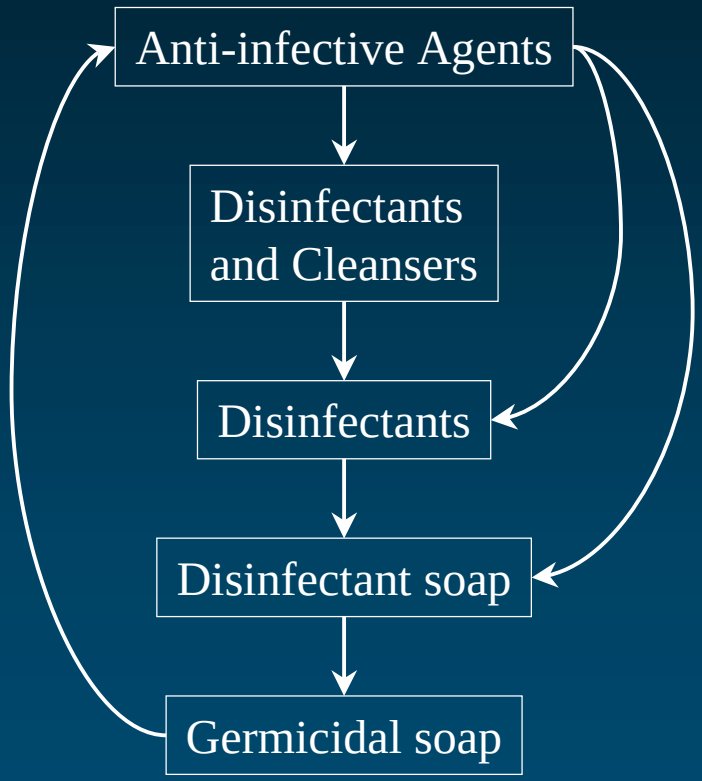
- Often task-driven
  rather than based on principles
- Usually suitable for information retrieval
  - Better recall
  - Precision may not be crucial
- Not necessarily suitable for reasoning
- But expected to be consistent structurally

AD in UMLS Contexts

- Multiple tree structures combined into a graph structure
- Directed acyclic graph (DAG)

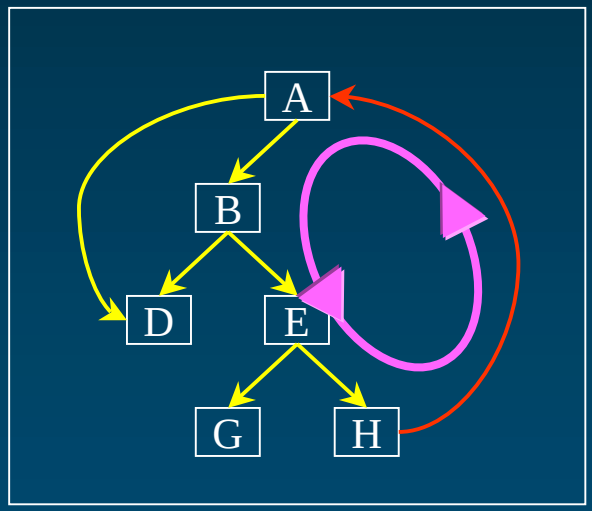


Actually, there are some cycles



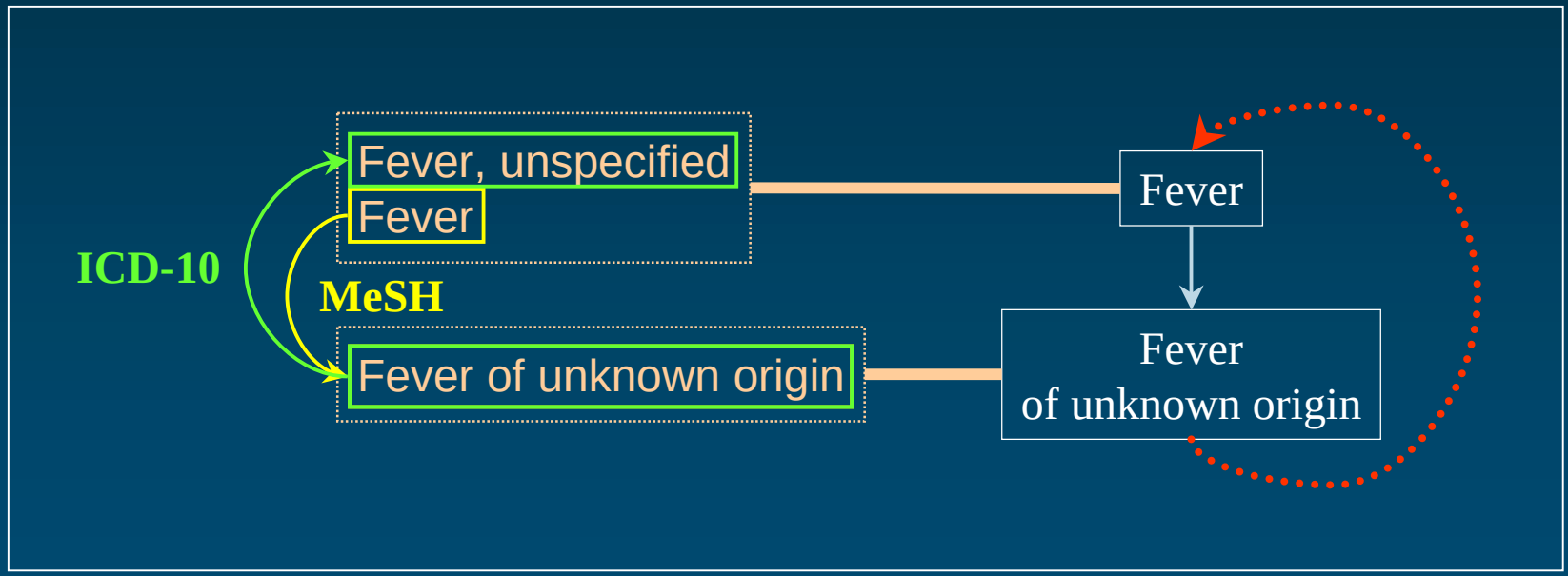
Issues with cycles

- Theoretical
  - Violate the antisymmetry property of partial ordering relations
- Practical
  - Loops in graph traversal
  - Impossible to perform transitive reduction



Cycle due to underspecification

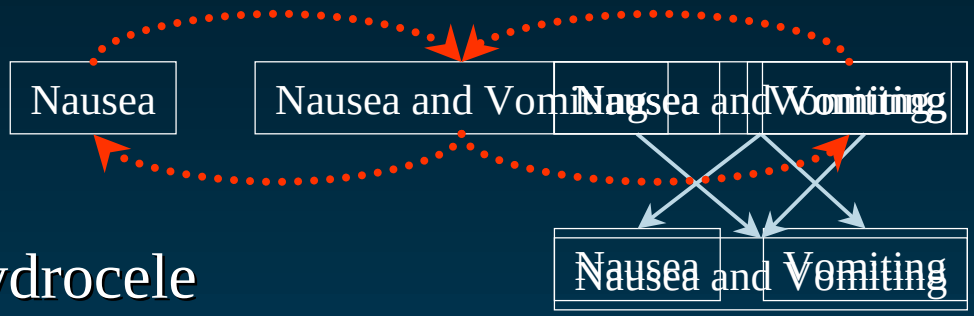
- Specified and underspecified terms
  - May appear at different levels in a source hierarchy
  - Are clustered into the same concept (same meaning)



Other causes

[Bodenreider, AMIA 2001]

- Compound terms
- Metadata
  - HYDROCELE, Hydrocele
- Classes and member
  - Purines, Purine
- Organizational conventions
  - Acid + Base ⇄ Salt + Water
- Idiopathic
  - Wrong relationships
  - Use of non-hierarchical relationships in “hierarchies”



Compensation mechanisms

Lexically-suggested hyponymic relations

Methods Overview

[Bodenreider & al., *TIA*, 2001]

- Syntactic analysis to identify adjectival modifiers
- Generate transformed terms by removing adjectival modifiers
- Map transformed terms to the UMLS
- Study the relationship between original term and transformed term in the UMLS, if any

Identify adjectival modifiers

- Underspecified syntactic analysis
  - Xerox part of speech tagger
  - SPECIALIST Lexicon (UMLS)
- Modifiers used: adjectives (+ adverbs)
- Modifiers identified in 64% of the terms
- Usually 1 to 2 modifiers
- Unique modifiers
  - 5400 adjectives
  - 69 adverbs

Transforming terms

- Remove any combination of modifiers found in the original term
- 2^n-1 transformed terms
  when the original term has n modifiers
- 104,000 transformed terms generated

acute infantile eczema

infantile eczema

acute eczema

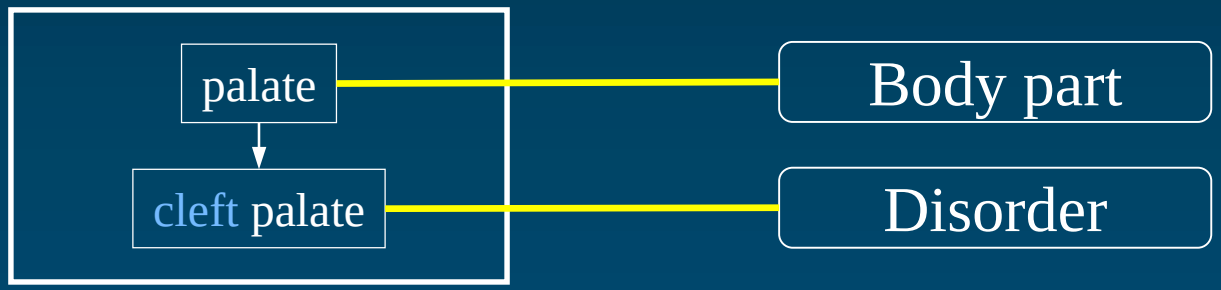
eczema

Mapping transformed terms to UMLS

- Increasing aggressiveness
  - Exact match
  - After normalization
- 25% of the transformed terms successfully mapped to UMLS

Excluding non-hyponymic relations

- If in hyponymic relation, original term and the transformed term should have the same semantic type (both Disease or both Procedure)
- Different semantic types in 10%



Checking relationships against UMLS

- Original term (OT) – Transformed term (TT)
  - Synonyms (same concept)
  - TT ancestor of OT
  - Siblings (inter-concept relationship)
  - Otherwise related

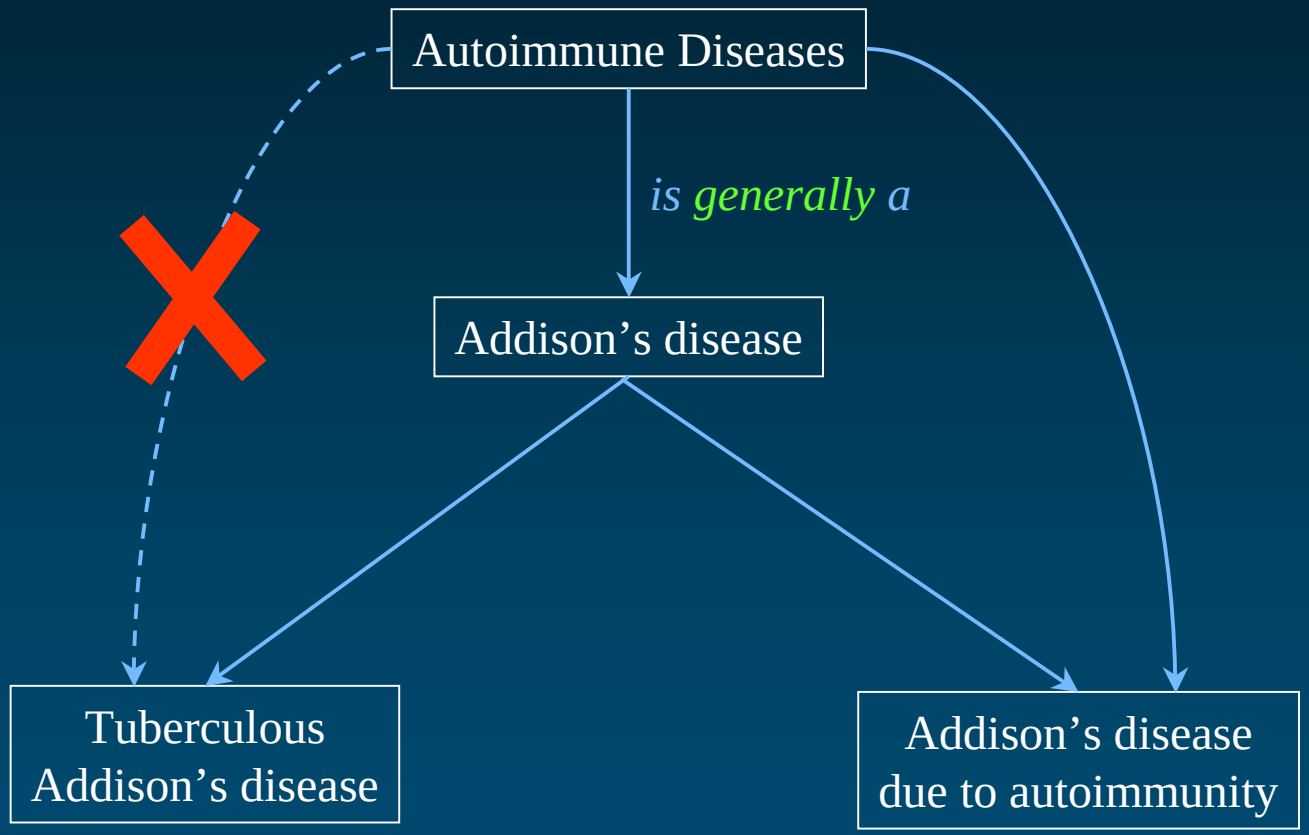
Lexically-suggested relationships / UMLS

- 28,851 pairs of terms
  - Original SNOMED term
  - Transformed term (found in UMLS)
- Corresponding relationship in the Metathesaurus
  - Hierarchical in 50% of the cases
  - « Sibling » in 25% of the cases
  - Missing in 25% of the cases

More limitations

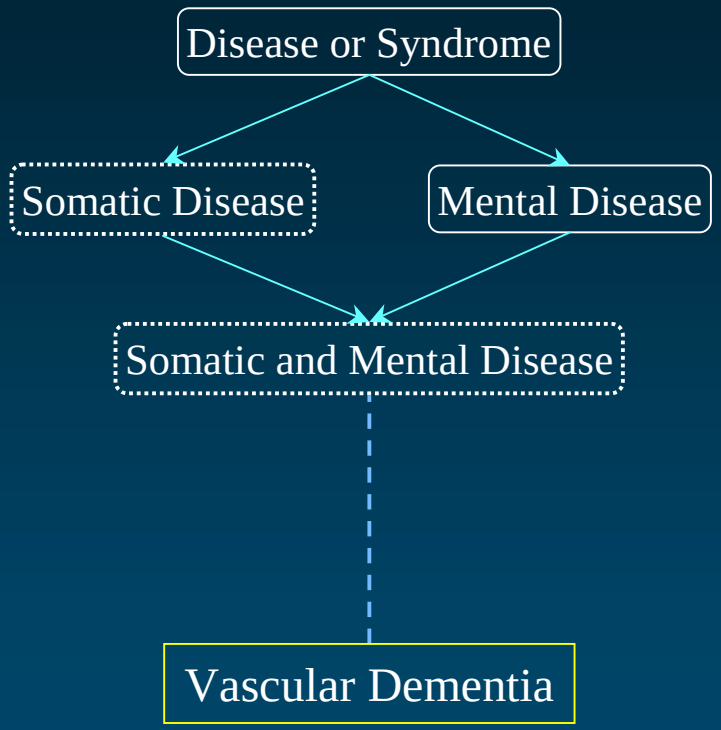
- Some missing / wrong relations are hard to detect
- Some relations are present but hard to find

Not all “isa” relationships are transitive!



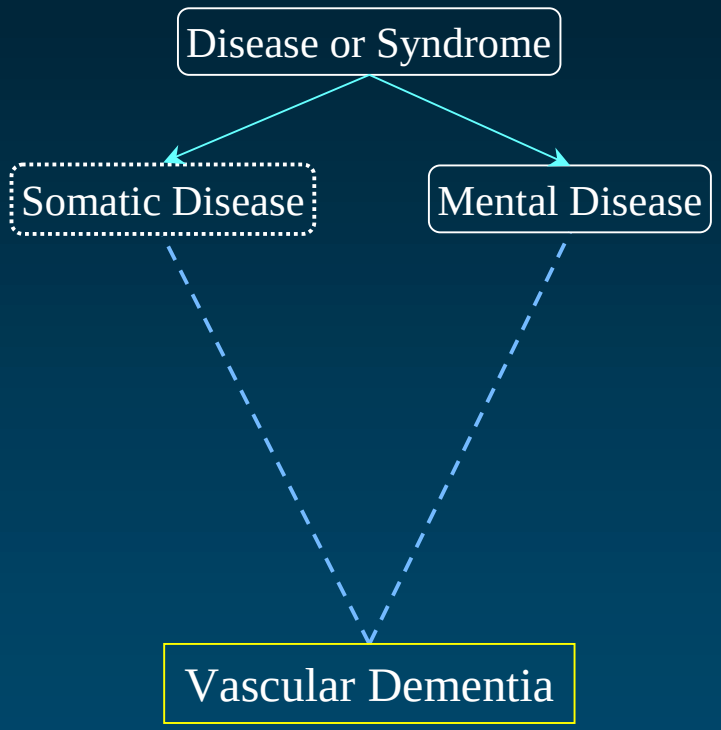
Semantic Network

[Burgun & al., *FOIS*, 2001]



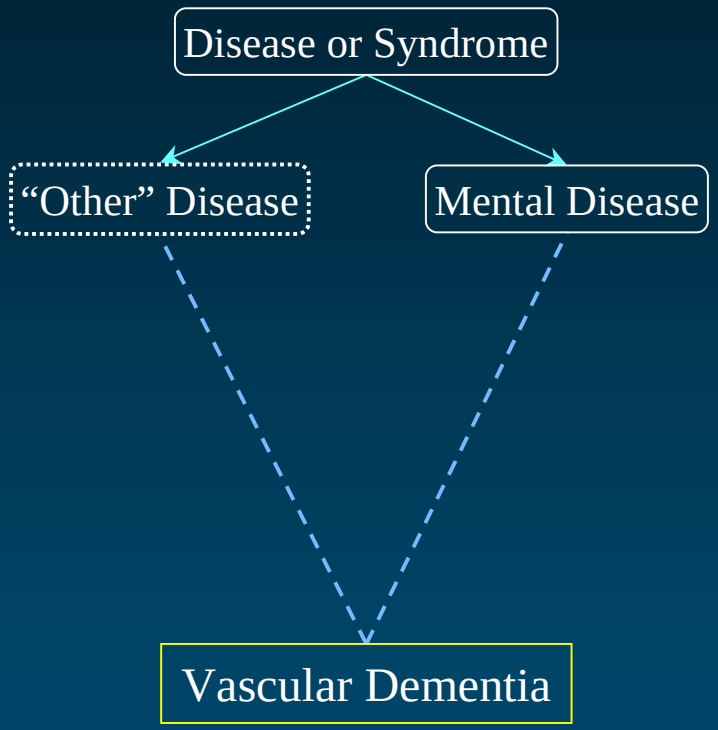
No hybrid type
Thus, multiple categorization

Semantic Network



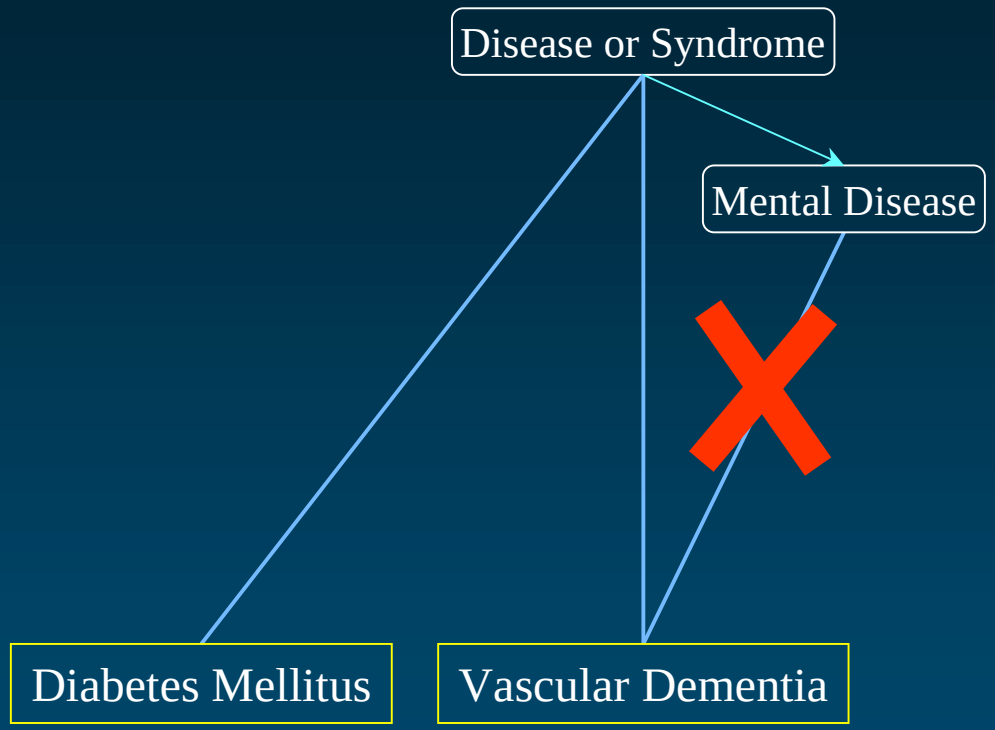
Ad hoc precision in hierarchies
Thus, no “Somatic disease” type

Semantic Network



No "other" type;
Assign to the common supertype

Semantic Network



Possible solutions

- Description logics
- Natural Language Processing
  (semantic interpretation of the terms)
- Comparing knowledge sources
  (alignment, inference)

Conclusions

Conclusions The up side

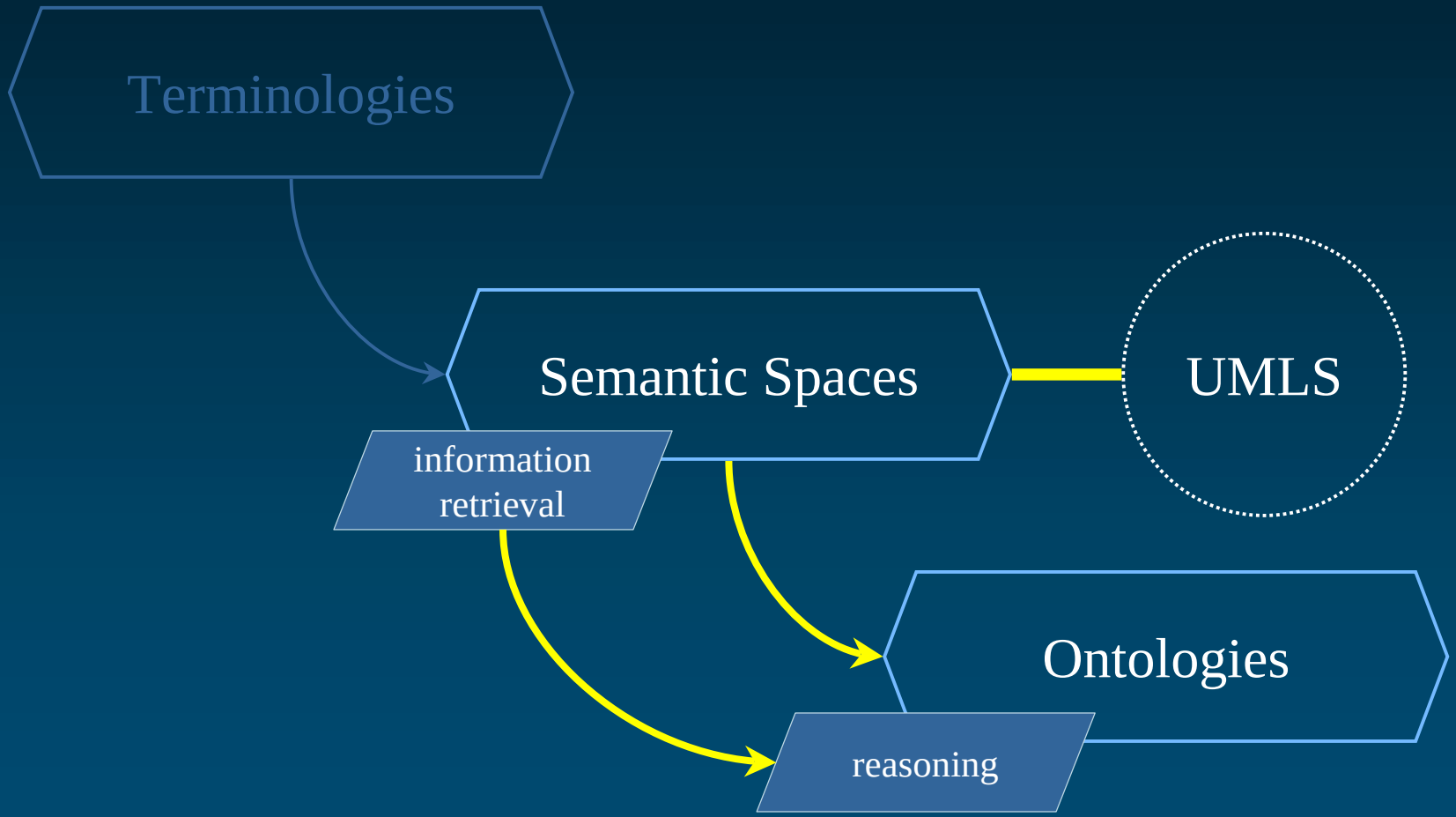
- Terminology integration is a step towards interoperability
  - Clusters of synonyms from different sources
  - Paths between terms from different sources

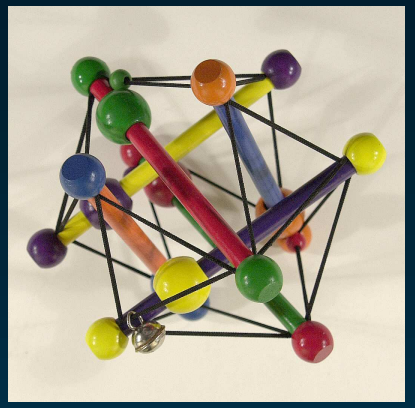
Conclusions The down side

- However, interoperability requires more than loosely aligned terminologies
- The UMLS does not claim to be an ontology
- The UMLS is, however, a resource for acquiring biomedical ontologies

Conclusions

Medical Ontology Research





Medical Ontology Research

Contact: olivier@nlm.nih.gov

Web: etbsun2.nlm.nih.gov:8000

NATIONAL LIBRARY OF MEDICINE

Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA